TRANSACTIONS

OF THE

AMERICAN MATHEMATICAL SOCIETY

EDITED BY

WILLIAM C. GRAUSTEIN

EINAR HILLE

C. C. MAC DUFFEE

WITH THE COÖPERATION OF

A. A. ALBERT	JESSE DOUGLAS	T. H. HILDEBRANDT
E. P. LANE	R. E. LANGER	SAUNDERS MACLANE
MARSTON MORSE	OYSTEIN ORE	H. L. RIETZ
H. P. ROBERTSON	M. H. STONE	J. L. SYNGE
GABOR SZEGÖ	G. T. WHYBURN	OSCAR ZARISKI

VOLUME 47

JANUARY TO JUNE, 1940

PUBLISHED BY THE SOCIETY

MENASHA, WIS., AND NEW YORK

1940

BOSTON UNIVERSITY
COLLEGE OF LIBERAL ARTS
LIBRARY

Chester C. Corbin Library Fund
Bound Dec. 1940

Composed, Printed and Bound by
The Collegiate Press
George Banta Publishing Company
Menasha, Wisconsin

64485

QA
1
F001
v.47

TABLE OF CONTENTS

VOLUME 47, JANUARY TO JUNE, 1940

ALEXANDER, H. W. The role of the mean curvature in the immersion theory of surfaces 230
AMBROSE, W. On measurable stochastic processes 66
BAER, R. Nilpotent groups and their generalizations 393
——— Nets and groups. II 435
BOCHNER, S. Harmonic surfaces in Riemann metric 146
DE CICCO, J. The geometry of fields of lineal elements 207
DOOB, J. L. Regularity properties of certain families of chance variables 455
DUNFORD, N., and PETTIS, B. J. Linear operations on summable functions 323
GRAUSTEIN, W. C. Harmonic minimal surfaces 173
HALL, D. W. On a decomposition of true cyclic elements 305
HESTENES, M. R., and MCSHANE, E. J. A theorem on quadratic forms and its application in the calculus of variations 501
HILLE, E. Contributions to the theory of Hermitian series. II. The representation problem 80
MCSHANE, E. J., and HESTENES, M. R. A theorem on quadratic forms and its application in the calculus of variations 501
PALL, G. On the arithmetic of quaternions 487
PETTIS, B. J., and DUNFORD, N. Linear operations on summable functions 323
PHILLIPS, R. S. Integration in a convex linear topological space 114
PRICE, G. B. The theory of integration 1
PUCKETT, W. T. On 0-regular surface transformations 95
SCHILLING, O. F. G. Regular normal extensions over complete fields 440
SCHWID, N. A correction to "The asymptotic forms of the Hermite and Weber functions" 322
SCOTT, W. T., and WALL, H. S. A convergence theorem for continued fractions 155
SZEGÖ, G. On the gradient of solid harmonic polynomials 51
WALL, H. S., and SCOTT, W. T. A convergence theorem for continued fractions 155

WALSH, J. L. On the degree of convergence of sequences of rational functions 254
——— Note on the degree of convergence of sequences of analytic functions 293

THE THEORY OF INTEGRATION*

BY

G. BALEY PRICE

1. **Introduction.** There are several types of integration for numerically-valued functions, one of which is associated with the ideas and methods of Cauchy, Riemann, and Lebesgue, another of which is connected with the method of Perron, and so on. This paper will be concerned with the integration of functions with values in a Banach space by methods in the Cauchy-Riemann-Lebesgue tradition.

Numerous other studies of integration in abstract spaces have been made in recent times. Noteworthy among these are (a) papers by Graves [1],† Bochner [1], Dunford [1], and Gowurin [1]; (b) a paper by Birkhoff [1]; (c) and other papers by Birkhoff [2], Dunford [2], and Pettis [1]. In spite of the fact that the theories of integration in both (a) and (b) are developments of the ideas of Cauchy, Riemann, and Lebesgue, there is an underlying unity of method in the four papers in (a) which groups them and distinguishes them from the paper in (b). The Lebesgue integral for numerically-valued functions can be developed by either of these two general methods, but for functions with values in a Banach space they are not equivalent.

The purpose of the present paper is to study the two general methods used in (a) and (b) for integrating functions with values in a Banach space, to determine their interrelations and limitations, and to generalize and extend them.

We proceed to describe the extensions and generalizations obtained.

(1) *Generalized convex sets.* One of the important contributions made by Birkhoff [1] was in showing the fundamental importance of convex sets in the theory of integration. In §2 further properties of convex sets are established. In §3 generalized convex sets are defined and their properties established. These sets form an extensive class which includes convex sets; they have all the properties of convex sets needed in integration.

(2) *Generalized measure functions.* By using the generalized convex sets described in (1), it is shown in §§8–19 that the values of the measure function in both Bochner's and Birkhoff's integral, which they assumed to be numbers, may be bounded transformations with certain restrictions. The use of the gen-

* Presented to the Society, December 28, 1937; received by the editors October 13, 1938, and, in revised form, July 31, 1939.

† Numbers in square brackets refer to the bibliography at the end.

BOSTON UNIVERSITY
COLLEGE OF LIBERAL ARTS
LIBRARY

eralized convex sets thus enables us to obtain an integral more general than that of Birkhoff.

(3) *Measurable functions.* A definition of measurable function is given in §14, and the properties of this class of functions are established in §15. It is shown in §19 that this class includes the measurable functions of Bochner as a proper subclass, and necessary and sufficient conditions that a measurable function be measurable in the sense of Bochner are established. In §16 a sufficient condition that a measurable function be integrable by the method of Birkhoff is established.

(4) *The Darboux theory of the upper and lower integral.* In Part V the properties of the Riemann-Stieltjes integral in abstract spaces are established. It is shown in §21 that there is a closed convex set which replaces the upper and lower integral for numerically-valued functions. In §25 it is shown that, at least in certain cases, the extreme points of this convex set (see Price [1]) are the exact analogues of the upper and lower integrals themselves.

(5) *A new class of Riemann-Stieltjes integrable functions.* In §§22–24 a new class of integrable functions and their integrals are studied. The values of these functions are sets instead of a single element in a Banach space; their integrals, of the Riemann-Stieltjes type, are closed convex sets. The integrals of these functions are additive in an extended sense (see Theorem 23.11).

(6) *A new class of measurable functions of the Bochner type.* It is shown in §26 that Bochner's method can be used to define measurable and summable functions whose values are sets in a Banach space; the integrals of these functions are closed convex sets and have the properties of the Lebesgue integral.

(7) *A new class of measurable functions of the Bochner-Dunford type.* It is shown in §27 that the method used by Bochner and Dunford can be extended still further to define measurable and summable functions whose values are sets in a Banach space; the integrals of these functions are closed convex sets and have the properties of the Lebesgue integral. This integral is the most general one obtained with the properties of the Lebesgue integral.

For numerically-valued functions there is only one Lebesgue integral—an integral which includes all others of the Cauchy-Riemann-Lebesgue type. There are many ways of developing the theory of this integral, but all methods lead to the same result. It is natural to ask whether the same is true for functions with values in a Banach space. The answer is in the negative. Birkhoff [1, p. 377] showed that his integral includes those of Graves and Bochner. But the integral of Bochner does not include that of Graves; an example given by Graves [1, p. 166] proves this fact. The integrals of Bochner and Dunford are known to be equivalent (see Dunford [3, p. 475]). The situa-

tion is changed radically, however, by the extensions given above. Neither the original Birkhoff integral nor the generalization of it described in (2) includes the generalized Bochner integral of the functions in (6); furthermore, neither the Bochner integral nor its extensions (see (2) and (6) above) include the generalized Bochner-Dunford integral of the functions described in (7). The generalization of the Bochner integral in (2) is the Gowurin integral; the generalized Birkhoff integral in (2) overlaps the Gowurin integral, but neither includes all cases of the other. It appears therefore that the extensions of the various methods associated with the Cauchy-Riemann-Lebesgue processes of integration lead not to one theory in abstract spaces, but to several.

As for methods and technique, there is one new element in the present case. The Hausdorff distance between sets in a Banach space is introduced (see §2); with this metric these sets become the elements of a complete metric space. In §4 the convergence of infinite series whose terms are elements of this metric space is studied; the types of convergence are numerous. Birkhoff also studied infinite series whose terms are sets in a Banach space, but without the Hausdorff metric. In §5 the interrelations of the various types of convergence introduced in §4 are examined, and in addition they are compared with Birkhoff's unconditional summation of complexes (see Birkhoff [1, pp. 361–364]). In §6 a theorem is proved which is fundamental in Birkhoff's method of integration; it is an extension of one given by Birkhoff [1, p. 364]. It is this theorem which requires the introduction of convex and generalized convex sets in Birkhoff's method of integration.

The use of the Hausdorff metric does not lead to a more general integral of the Birkhoff type (see Theorem 12.2), but it does make possible the generalizations listed under (4), (5), (6), (7) above. Furthermore, its use seems to simplify the details of some proofs, and to permit all of them to be stated in more familiar form and language.

The paper closes with applications of some of the results on integration to Fourier series and singular integrals. It is shown that the limit of the Fejér integral under very general conditions represents the value of the function—even when the values of the function are sets and it is impossible to establish the usual connection between the Fourier series and the Fejér integral. It is shown that the Fourier series of a function of bounded total variation converges to the usual value. It is necessary to construct a new proof of this theorem since the mean value theorems usually used are lacking.

Part I. Generalized convex sets

2. **The algebra of complexes.** We shall first explain the notation to be used. Let $\mathfrak{B}$ be a Banach space, that is, a space of type (B) (see Banach [1,

p. 53]), with elements $f, g, \cdots$ whose norms are $\|f\|, \|g\|, \cdots$. Let F denote a set or complex of elements f and aF the complex of elements af when a is any real number. If R is any set of real numbers r, let RF denote the set of elements rf, $r \in R, f \in F$. By the sum $F_1 + \cdots + F_n$, or $\sum_1^n F_i$, is meant the complex of elements $f_1 + \cdots + f_n, f_i \in F_i$. The closure of F is $\overline{F}$. The convex hull of F is $C[F]$, and the closed convex hull is $\overline{C}[F]$. The closure of $F+G$ will be represented also by $F+^*G$. Further details will be found in Birkhoff [1, pp. 358–365].

Throughout the paper a set $F \subset \mathfrak{B}$ is understood to be bounded unless there is a statement to the contrary.

(2.1) DEFINITION. *Let F_1, F_2 be two sets in $\mathfrak{B}$. Let d_1 be the upper bound of distances from points of F_1 to F_2 and d_2 the upper bound of distances from points of F_2 to F_1. The larger of the numbers d_1, d_2 is called the distance $D(F_1, F_2)$ between F_1 and F_2.*

This definition of distance between two sets in a metric space was introduced by Hausdorff [1, pp. 145–146]; it has been used by Blaschke [1, pp. 59–66].

The diameter $D(F-F, 0)$ of a set F is denoted by $\rho(F)$.

The following relations are consequences of the above definitions:

$$(2.2) \qquad D(F_1, F_2) = D(F_2, F_1);$$

$$(2.3) \qquad D(F_1, F_2) \geqq 0;$$

$$(2.4) \qquad D(F_1, F_2) = 0 \text{ equivalent to } \overline{F}_1 = \overline{F}_2;$$

$$(2.5) \qquad D(F_1, F_3) \leqq D(F_1, F_2) + D(F_2, F_3);$$

$$(2.6) \qquad D(aF_1, aF_2) = |a| D(F_1, F_2) \qquad \text{for every real number } a;$$

$$(2.7) \qquad D(F_1, F_2) = D(F_1, \overline{F}_2);$$

$$(2.8) \qquad D(F_1 + \cdots + F_n, G_1 + \cdots + G_n) \leqq D(F_1, G_1) + \cdots + D(F_n, G_n);$$

$$(2.9) \qquad D(C[F_1], C[F_2]) \leqq D(F_1, F_2);$$

$$(2.10) \qquad D(F + G_1, F + G_2) \leqq D(G_1, G_2);$$

$$(2.11) \qquad \rho(F_i) \leqq \rho(F_1 + F_2) \leqq \rho(F_1) + \rho(F_2) \qquad (i = 1, 2);$$

$$(2.12) \qquad \rho(C[F]) = \rho(F), \qquad D(C[F], 0) = D(F, 0).$$

The first six of these relations follow at once from Definition 2.1; (2.11) and (2.12) were given by Birkhoff [1, pp. 368, 360]. Furthermore, (2.10) follows at once from (2.8) and (2.4). It can be shown by means of examples that the inequality may hold in (2.10). We shall give proofs of (2.8) and (2.9).

Proof of (2.8). Let $\epsilon > 0$ be given. Then it is possible to choose either a point $(f_1 + \cdots + f_n) \in (F_1 + \cdots + F_n)$ so that

$$d(f_1 + \cdots + f_n, G_1 + \cdots + G_n) \geqq D(F_1 + \cdots + F_n, G_1 + \cdots + G_n) - \epsilon$$

or a point $(g_1 + \cdots + g_n) \in (G_1 + \cdots + G_n)$ so that

$$d(g_1 + \cdots + g_n, F_1 + \cdots + F_n) \geqq D(F_1 + \cdots + F_n, G_1 + \cdots + G_n) - \epsilon,$$

where $d(f, F)$ represents the distance from f to the set F. It follows from Definition 2.1 that this choice is possible. Without loss of generality we may assume that the first case occurs. Then we can hold $f_1, \cdots, f_n$ fixed and choose points $g_1 \in G_1, \cdots, g_n \in G_n$ so that

$$\|f_1 - g_1\| < D(F_1, G_1) + \epsilon/n, \cdots, \|f_n - g_n\| < D(F_n, G_n) + \epsilon/n.$$

Then

$$\begin{aligned}\|(f_1 + \cdots + f_n) - (g_1 + \cdots + g_n)\| &\leqq \|f_1 - g_1\| + \cdots + \|f_n - g_n\| \\ &\leqq D(F_1, G_1) + \cdots + D(F_n, G_n) + \epsilon.\end{aligned}$$

But since

$$\begin{aligned}D(F_1 + \cdots + F_n, G_1 + \cdots + G_n) - \epsilon &\leqq d(f_1 + \cdots + f_n, G_1 + \cdots + G_n) \\ &\leqq \|(f_1 + \cdots + f_n) - (g_1 + \cdots + g_n)\|,\end{aligned}$$

we see that

$$D(F_1 + \cdots + F_n, G_1 + \cdots + G_n) - \epsilon \leqq D(F_1, G_1) + \cdots + D(F_n, G_n) + \epsilon.$$

Since ϵ is arbitrary, (2.8) follows and the proof is complete.

Proof of (2.9). The distance of a point of F_2 from $C[F_1]$ does not exceed its distance from F_1 since $F_1 \subseteqq C[F_1]$. Then the parallel set to $C[F_1]$ at the distance $D(F_1, F_2)$, that is, the set of all those points at distance from $C[F_1]$ equal to or less than $D(F_1, F_2)$, is a convex set which contains F_2. Then the distance from any point of $C[F_2]$ to $C[F_1]$ is equal to or less than $D(F_1, F_2)$. In the same way we show that the distance from any point of $C[F_1]$ to $C[F_2]$ is equal to or less than $D(F_1, F_2)$. Then (2.9) follows from Definition 2.1, and the proof is complete.

It is possible to give an example in which $D(C[F_1], C[F_2]) < D(F_1, F_2)$.

With the distance introduced in Definition 2.1, the complexes F are the elements of a complete metric space (see Price [3]). Two complexes F_1, F_2 are metrically identical when their closures coincide in the point-set sense.

(2.13) THEOREM. *If $F_1, \cdots, F_n$ are convex sets and $a_1, \cdots, a_n$ are any real numbers, then $\sum_1^n a_i F_i$ is a convex set, and*

$$\sum_1^n a_i \overline{F}_i \subseteqq \overline{\sum_1^n a_i F_i}.$$

This theorem was given by Birkhoff [1, pp. 359–360].

(2.14) THEOREM. *If $F_1, \cdots, F_n$, $F_i \subset \mathfrak{B}$, are sets which are compact in $\mathfrak{B}$, and if $a_1, \cdots, a_n$ are any real numbers, then*

$$\sum_1^n a_i \overline{F}_i = \overline{\sum_1^n a_i F_i}.$$

The proof is omitted.

(2.15) THEOREM. *Let $F_1, \cdots, F_n$, $F_i \subset \mathfrak{B}$, be convex sets. If g is any extreme point of the convex set $\sum_1^n a_i F_i$, then $g = \sum_1^n a_i g_i$, and $g_1, \cdots, g_n$ are extreme points of $F_1, \cdots, F_n$ respectively.*

A point of a convex set is an extreme point if it is not an interior point of any segment which belongs to the set (see Price [1, p. 57]). Since g is a point of $\sum_1^n a_i F_i$, there exist points $g_1, \cdots, g_n$, $g_i \in F_i$, such that $g = \sum_1^n a_i g_i$. Suppose that one of the points $g_1, \cdots, g_n$, say g_1, is not an extreme point. Then there exist two points g_1', g_1'' in F_1 such that $g_1 = \theta g_1' + (1-\theta) g_1''$, $0 < \theta < 1$. Then $g = \theta g' + (1-\theta) g''$, where $g' = a_1 g_1' + a_2 g_2 + \cdots + a_n g_n$, $g'' = a_1 g_1'' + a_2 g_2 + \cdots + a_n g_n$, and g is not an extreme point. This contradiction establishes the theorem.

(2.16) THEOREM. *If R is a closed interval of real numbers, and if $F \subset \mathfrak{B}$ is closed, then RF is closed. If in addition R does not contain the number zero in its interior, and if F is convex, then RF is convex also.*

(2.17) COROLLARY. *If R, R' are arbitrary closed intervals of real numbers, and if F is the set $R'f$, f a fixed element in $\mathfrak{B}$, then RF is closed and convex.*

The conditions of neither the theorem nor the corollary are necessary.

(2.18) THEOREM. *If R is any set of numbers such that zero is not an interior point of its closed convex hull $\overline{C}[R]$, and if F is any set in $\mathfrak{B}$, then $\overline{C}[RF] = \overline{C}[R]\overline{C}[F]$.*

If $r_0 \in R$, then $r_0 F \subseteq \overline{C}[RF]$ because of the convexity properties of $\overline{C}[RF]$. Again, if r_1 is a limit point of points of R, then $r_1 F \subseteq r_1 \overline{C}[F] \subseteq \overline{C}[RF]$ because $\overline{C}[RF]$ is closed. Finally, let $\overline{C}[R] \equiv (\alpha \leq r \leq \beta)$. Then by the statements just made, $\alpha \overline{C}[F] \subseteq \overline{C}[RF]$, $\beta \overline{C}[F] \subseteq C[RF]$; hence, because of the convexity once more, $\overline{C}[R]\overline{C}[F] \subseteq \overline{C}[RF]$. Also $RF \subseteq \overline{C}[R]\overline{C}[F]$ and, by Theorem 2.16, $\overline{C}[R]\overline{C}[F]$ is closed and convex. But $\overline{C}[RF]$ has no proper subset with these properties, and therefore $\overline{C}[RF] = \overline{C}[R]\overline{C}[F]$. The proof is complete.

(2.19) COROLLARY. *Let R_1, R_2 be two sets of numbers such that zero is not an interior point of $\overline{C}[R_1]$, $\overline{C}[R_2]$, and let F_1, F_2 be two arbitrary sets in $\mathfrak{B}$. Then*

$$\overline{C}[R_1F_1 + R_2F_2] = \overline{C}[R_1]C[F_1] + \overline{C}[R_2]C[F_2]. \tag{2.20}$$

This corollary follows at once from the theorem and $C[R_1F_1+R_2F_2] = C[R_1F_1]+C[R_2F_2]$ (see Birkhoff [1, p. 360]).

(2.21) THEOREM. *Let R_1, R_2 be two sets of numbers such that zero is not an interior point of $C[R_1]$, $C[R_2]$, and let F be any set in $\mathfrak{B}$. Then*

$$\overline{C}[R_1]\overline{C}[F] + \overline{C}[R_2]\overline{C}[F] = \overline{C}[R_1 + R_2]\overline{C}[F]; \tag{2.22}$$

$$D(\overline{C}[R_1]\overline{C}[F], \overline{C}[R_2]\overline{C}[F]) \leqq D(\overline{C}[R_1], \overline{C}[R_2])D(\overline{C}[F], 0). \tag{2.23}$$

Proof of (2.23). Consider the distance from any point $r_1 f \in \overline{C}[R_1]\overline{C}[F]$ to $\overline{C}[R_2]\overline{C}[F]$. We can find a point $r_2 \in \overline{C}[R_2]$ such that $|r_1 - r_2| \leqq D(\overline{C}[R_1], \overline{C}[R_2])$. Then the distance from $r_1 f$ to $\overline{C}[R_2]\overline{C}[F]$ does not exceed

$$\|r_1 f - r_2 f\| \leqq |r_1 - r_2| \cdot \|f\| \leqq D(\overline{C}[R_1], \overline{C}[R_2])D(\overline{C}[F], 0).$$

In the same way it can be shown that the distance from points of $\overline{C}[R_2]\overline{C}[F]$ to $\overline{C}[R_1]\overline{C}[F]$ does not exceed the same number. The proof is then complete.

3. **Generalized convex sets.** In this section we shall establish the fundamental properties of a linear operator whose domain and range are sets $F \subset \mathfrak{B}$.

Let $\mathfrak{T}$ be the normed linear space of linear transformations T whose domain is $\mathfrak{B}$ and whose range is in $\mathfrak{B}$. Thus if $T \in \mathfrak{T}$ and a, b are real numbers, we have $Tf \in \mathfrak{B}$, $T(af+bg) = aTf+bTg$, $\|Tf\| \leqq \|T\| \cdot \|f\|$. If T_1, $T_2 \in \mathfrak{T}$, then T_1T_2 and T_1+T_2 also belong to $\mathfrak{T}$. The identical transformation I belongs to $\mathfrak{T}$.

Let t denote a set of elements of $\mathfrak{T}$: $t \equiv (T_1, \cdots, T_r)$. Here $T_i \in \mathfrak{T}$, and $r = r(t)$ is a positive integer.

(3.1) DEFINITION. *The product t_1t_2 of t_1 and t_2, where $t_1 \equiv (T_{1i})$, $i = 1, \cdots, r_1$, and $t_2 \equiv (T_{2j})$, $j = 1, \cdots, r_2$, is $t \equiv (T_{1i}T_{2j})$, $i = 1, \cdots, r_1$ and $j = 1, \cdots, r_2$.*

Let C^* denote a set of elements t which satisfies the following hypotheses:

(3.2) HYPOTHESIS. *$t \in C^*$ implies $T_1 + \cdots + T_r = I$.*

(3.3) HYPOTHESIS. *$t_1, t_2 \in C^*$ implies $t_1t_2 \in C^*$.*

(3.4) HYPOTHESIS. *There exists a constant W, depending only on C^*, such that for every $t \in C^*$ and any set of points $f_1, \cdots, f_r$ in $\mathfrak{B}$*

$$\left\| \sum_1^r T_i f_i \right\| \leqq W \max_i \|f_i\|. \tag{3.5}$$

When Hypothesis 3.4 is satisfied, we shall assume that W represents the lower bound of the constants for which (3.5) holds. The W-property of a set of transformations t was first studied by Gowurin [1]. We shall say C^* is of bounded variation if there exists a constant V such that for every $t \in C^*$,

$\sum_1^r \|t_i\| \leqq V$. Hypothesis 3.5 is weaker than bounded variation and is implied by it. The convex operator C has $V=1$ and $W=1$.

It should be remarked that if t_1, t_2 satisfy Hypothesis 3.2, then t_1t_2 automatically satisfies the same hypothesis.

We shall use C^* also to denote an operator as explained in the following definition.

(3.6) DEFINITION. *The transform of a set $F \subset \mathfrak{B}$ by C^*, denoted by $C^*[F]$, is the set of all points $\sum_1^r T_i f_i$, where $f_i \in F$ and $t \equiv (T_1, \cdots, T_r) \in C^*$. If F is the null set, $C^*[F]$ is the null set.*

The fundamental properties of this operator are given in the following theorems and corollaries.

(3.7) THEOREM. $C^*[F+G]=C^*[F]+C^*[G]$.

(3.8) COROLLARY. *Let T_1, T_2 be any two numbers, or more generally, any two elements of $\mathfrak{T}$ which commute with all the transformations T_i in $t \in C^*$; then* $C^*[T_1F+T_2G]=T_1C^*[F]+T_2C^*[G]$.

(3.9) THEOREM. *$F \leqq G$ implies* $C^*[F] \leqq C^*[G]$.

(3.10) THEOREM. $F \leqq C^*[F]$.

(3.11) THEOREM. $C^*[\prod_{(\nu)} C^*[F_\nu]]=\prod_{(\nu)} C^*[F_\nu]$.

(3.12) COROLLARY. $C^*[C^*[F]]=C^*[F]$.

(3.13) THEOREM. *$t \equiv (T_1, \cdots, T_r) \in C^*$ implies* $C^*[F]=T_1C^*[F]+\cdots+T_rC^*[F]$.

(3.14) THEOREM. $D(C^*[F_1], C^*[F_2]) \leqq WD(F_1, F_2)$.

(3.15) THEOREM. $D(F, 0) \leqq D(C^*[F], 0) \leqq WD(F, 0)$, *and also* $D(F, 0) \leqq D(C^*[F], 0) \leqq D(F, 0)+W\rho(F)$.

(3.16) THEOREM. $\rho(F) \leqq \rho(C^*[F]) \leqq W\rho(F)$.

From (3.8), (3.9), and (3.15) it follows that C^* is a homogeneous, additive, monotone, and bounded operator. It should be remarked that the proofs of (3.7)–(3.13) do not depend on Hypothesis 3.4; only the proofs of (3.14)–(3.16) depend on this third hypothesis. Theorems 3.14–3.16 are generalizations of properties of the convex operator C already given in (2.9) and (2.12); the latter can be obtained from the former by setting $W=1$.

Proof of (3.7). Since the transformations T_i in $t \in C^*$ are linear by hypothesis, it is clear that $C^*[F+G] \leqq C^*[F]+C^*[G]$. Thus we have only to show that $C^*[F]+C^*[G] \leqq C^*[F+G]$.

Let $\sum_1^{r_1} T_{1i} f_i$, $\sum_1^{r_2} T_{2j} g_j$ be any two points in $C^*[F]$, $C^*[G]$ respectively. The

proof will be completed by showing that $\sum_1^{r_1} T_{1i}f_i + \sum_1^{r_2} T_{2j}g_j$, an arbitrary point of $C^*[F]+C^*[G]$, is identical with

$$\sum_{i=1}^{r_1}\sum_{j=1}^{r_2} T_{1i}T_{2j}(f_i + g_j), \tag{3.17}$$

which is a point of $C^*[F+G]$ by Hypothesis 3.3. The proof follows from the fact that

$$\begin{aligned}\sum_{i=1}^{r_1} T_{1i}\sum_{j=1}^{r_2}[T_{2j}f_i + T_{2j}g_j] &= \sum_{i=1}^{r_1} T_{1i}\left[\sum_{j=1}^{r_2} T_{2j}f_i + \sum_{j=1}^{r_2} T_{2j}g_j\right] \\ &= \sum_{i=1}^{r_1} T_{1i}\left[f_i + \sum_{j=1}^{r_2} T_{2j}g_j\right] \\ &= \sum_{i=1}^{r_1} T_{1i}f_i + \sum_{j=1}^{r_2} T_{2j}g_j,\end{aligned}$$

where (3.2) has been used in the transformations. The proof is complete.

The proofs of the remaining theorems and corollaries are left to the reader.

A trivial example of an operator C^* is formed by the set of elements t where t is a set of positive rational fractions with even denominators whose sum is 1. A more interesting example is the following one in the space of square matrices of order 2. The norm of a matrix is the square root of the sum of the squares of its elements. Let t be the r matrices

$$\begin{Vmatrix} \alpha_i & 0 \\ 0 & \beta_i \end{Vmatrix} \qquad (i = 1, \cdots, r),$$

where $0 \leqq \alpha_i, \beta_i \leqq 1, \sum_i \alpha_i = \sum_i \beta_i = 1$. A transformation T is here matrix multiplication. The set of all such elements t forms an operator C^* which satisfies Hypotheses 3.2, 3.3, 3.4. This operator is of bounded variation with $V=2$. It therefore has the W-property, and it can be shown that $2^{1/2} \leqq W \leqq 2$. A similar operator C^* exists in the space of square matrices of order n; in this case $V=n$ and $n^{1/2} \leqq W \leqq n$.

It is necessary to observe also that there are methods for constructing operators C^*. Let $\{\tau\}$ be a class of transformations with the following properties:

(3.18) HYPOTHESIS. *$\tau \in \{\tau\}$ implies $\tau \in \mathfrak{T}$; $\tau \in \{\tau\}$ and $\tau \neq 0$ imply that τ has an inverse $\tau^{-1} \in \mathfrak{T}$; $\tau_1, \cdots, \tau_r \in \{\tau\}$ and $(\tau_1 + \cdots + \tau_r) \neq 0$ imply that $(\tau_1 + \cdots + \tau_r)$ has an inverse $(\tau_1 + \cdots + \tau_r)^{-1} \in \mathfrak{T}$.*

Let $\tau_1, \cdots, \tau_r$ be any r elements in $\{\tau\}$; let t, t' be $[(\tau_1 + \cdots + \tau_r)^{-1}\tau_1, \cdots, (\tau_1 + \cdots + \tau_r)^{-1}\tau_r]$, $[\tau_1(\tau_1 + \cdots + \tau_r)^{-1}, \cdots, \tau_r(\tau_1 + \cdots + \tau_r)^{-1}]$ respectively and C_0^* the set of all such t and t'. Then C_0^* satisfies Hypothesis 3.2.

In the usual way we can form the smallest set C^* which contains C_0^* and is closed with respect to the multiplication defined in (3.1). Then C^* satisfies Hypotheses 3.2, 3.3 and is a homogeneous, additive, and monotone operator.

(3.19) THEOREM. *Let C^* be the operator formed from $\{\tau\}$ as described, and let $\tau_1, \cdots, \tau_r$ be any elements in $\{\tau\}$. Then for any set $F \subset \mathfrak{B}$*

$$(\tau_1 + \cdots + \tau_r)C^*[F] = \tau_1 C^*[F] + \cdots + \tau_r C^*[F].$$

From Theorem 3.13 we have

$$C^*[F] = (\tau_1 + \cdots + \tau_r)^{-1}\tau_1 C^*[F] + \cdots + (\tau_1 + \cdots + \tau_r)^{-1}\tau_r C^*[F],$$

from which the theorem follows.

In the future C^* will be used to denote an operator which satisfies Hypotheses 3.2, 3.3, 3.4 and is therefore additive, homogeneous, monotone, and bounded.

(3.20) DEFINITION. *A set $F \subset \mathfrak{B}$ such that $F = C^*[F]$ will be called convex C^*.*

It follows from Corollary 3.12 that every set $C^*[F]$ is convex C^*.

PART II. INFINITE SERIES

4. **Convergence of infinite series.** We shall now consider various types of convergence of the infinite series

$$\sum_{k=1}^{\infty} f_k, \qquad f_k \in \mathfrak{B}; \tag{4.1}$$

$$\sum_{k=1}^{\infty} F_k, \qquad F_k \subset \mathfrak{B}. \tag{4.2}$$

As stated in §2 the space with elements $F \subset \mathfrak{B}$ is a *complete* metric space.

(4.3) DEFINITION. *The series* (4.1) *converges unconditionally to $f \in \mathfrak{B}$ if and only if every rearrangement α of all the terms of the series gives a series $\sum_1^\infty f_{\alpha(k)}$ which converges to f.*

(4.4) DEFINITION. *The series* (4.2) *converges strongly unconditionally to the sum $F \subset \mathfrak{B}$ if and only if every series $\sum_1^\infty f_k$, $f_k \in F_k$, converges unconditionally and F is the locus of the sums of these series.*

These two definitions have been given by Birkhoff [1, pp. 361–362].

(4.5) DEFINITION. *The series* (4.2) *converges normally if and only if the series $\sum_1^\infty D(F_k, 0)$ converges.*

(4.6) DEFINITION. *The series* (4.2) *converges regularly if and only if for each $\epsilon > 0$ there exists an $N = N(\epsilon)$ such that $D(\sum_{k=N}^{N+p} F_k, 0) < \epsilon$, $p \geqq 0$.*

It can be shown that this definition is equivalent to the following one.

(4.6′) DEFINITION. *The series* (4.2) *converges regularly if and only if for each* $\epsilon>0$ *there exists an* $M=M(\epsilon)$ *such that* $D(\sum_m^n F_k, 0)<\epsilon$, $n\geqq m\geqq M$.

(4.7) DEFINITION. *The series* $\sum_1^\infty F_k$ *converges and has the sum* $F\subset\mathfrak{B}$ *if and only if for each* $\epsilon>0$ *there exists an* $N=N(\epsilon)$ *such that* $D(\sum_1^n F_k, F)<\epsilon$, $n\geqq N$.

It follows from (2.7) that the sum of a convergent series is not uniquely determined. We agree, however, that the sum of a convergent series shall be taken as a closed set; with this convention the sum is unique.

Since we have associated two sums with the series $\sum_1^\infty F_k$, it will be convenient to introduce notation for them. If the series converges in the sense of Definition 4.7, its sum will be denoted by $\sum_1^\infty F_k$. If $\sum_1^\infty F_k$ converges strongly unconditionally in the sense of Definition 4.4, we indicate its (strong) sum by $S\sum_1^\infty F_k$.

(4.8) DEFINITION. *The series* $\sum_1^\infty F_k$ *converges weakly unconditionally to* $F\subset\mathfrak{B}$ *if and only if every rearrangement* α *of all the terms of the series gives a series* $\sum_1^\infty F_{\alpha(k)}$ *which converges to* F.

(4.9) THEOREM. *A necessary and sufficient condition that* $\sum_1^\infty f_k$ *be unconditionally convergent to* f *is that to every* $\epsilon>0$ *there correspond an integer* N *such that the sum* η *of any finite set of terms of* $\sum_1^\infty f_k$ *including* $f_1, \cdots, f_N$ *satisfies* $\|\eta-f\|<\epsilon$.

(4.10) THEOREM. *A necessary and sufficient condition that* $\sum_1^\infty f_k$ *be unconditionally convergent to* f *is that for each* $\epsilon>0$ *there exist an* M *such that* $M<k(1)<k(2)<\cdots<k(r)$ *implies*

$$\|f_{k(1)}+f_{k(2)}+\cdots+f_{k(r)}\|<\epsilon.$$

(4.11) THEOREM. *A necessary and sufficient condition that* $\sum_1^\infty F_k$ *converge strongly unconditionally is that to each* $\epsilon>0$ *there correspond an* M *such that* $M<k(1)<k(2)<\cdots<k(r)$ *implies*

$$D(F_{k(1)}+F_{k(2)}+\cdots+F_{k(r)}, 0)<\epsilon.$$

These theorems have been given by Birkhoff [1, pp. 361–362].

(4.12) THEOREM. *A necessary and sufficient condition that* $\sum_1^\infty F_k$ *converge* (*see Definition* 4.7) *is that for each* $\epsilon>0$ *there exist an* N *such that* $D(\sum_{k=1}^{N+p} F_k, \sum_{k=1}^{N} F_k)<\epsilon$ *for all* $p\geqq 1$.

The proof follows easily from properties of the distance function given in §2. The proof that the condition is sufficient depends on the fact that the space with elements $F\subset\mathfrak{B}$ is a complete metric space.

(4.13) THEOREM. *A necessary and sufficient condition that* (4.2) *be weakly unconditionally convergent to F is that to every* $\epsilon>0$ *there correspond a number N such that the sum* Φ *of any finite set of terms of* (4.2) *including* $F_1, \cdots, F_N$ *satisfies* $D(\Phi, F)<\epsilon$.

This theorem is entirely similar to Theorem 4.9.

(4.14) THEOREM. *A sufficient condition that* $\sum_1^\infty F_k$ *be weakly unconditionally convergent to F is that for each* $\epsilon>0$ *there exist an M such that* $M<k(1)<k(2)<\cdots<k(r)$ *implies*

$$D(F_{k(1)}+F_{k(2)}+\cdots+F_{k(r)}, 0)<\epsilon.$$

We have

$$D\Big(\sum_{k=1}^m F_k, \sum_{k=1}^n F_k\Big) = D\Big(\sum_{k=1}^n F_k + \sum_{k=n+1}^m F_k, \sum_{k=1}^n F_k\Big) \leqq D\Big(\sum_{k=n+1}^m F_k, 0\Big) < \epsilon$$

provided only that $m\geqq n\geqq M$. From Theorem 4.12 it follows that $\sum_1^\infty F_k$ converges and has a sum F; then $D(\sum_1^m F_k, F)\leqq\epsilon$, $m\geqq M$. Now consider any rearrangement $\sum_1^\infty F_{\alpha(k)}$ of $\sum_1^\infty F_k$. We have

$$D\Big(\sum_{k=1}^\nu F_{\alpha(k)}, F\Big) \leqq D\Big(\sum_{k=1}^\nu F_{\alpha(k)}, \sum_{k=1}^M F_k\Big) + D\Big(\sum_{k=1}^M F_k, F\Big) < \epsilon+\epsilon$$

for all ν which are so large that $\sum_1^\nu F_{\alpha(k)}$ contains $F_1, \cdots, F_M$. Thus the series converges and has the sum F. The proof is complete.

If each set F_k in (4.2) consists of a single element f_k, weak unconditional convergence and strong unconditional convergence are identical and identical with unconditional convergence of (4.1) as defined in (4.3).

(4.15) THEOREM. *If* $\sum_1^\infty F_k$ *is strongly unconditionally convergent, and if* F_k *is convex* C^* *for all k, then the sum* $S\sum_1^\infty F_k$ *is convex* C^*.

(4.16) THEOREM. *If* $\sum_1^\infty F_k$ *converges, and if* F_k *is convex* C^* *for all k, then the sum* $\sum_1^\infty F_k$ *is convex* C^*.

We have

$$\begin{aligned} D\Big(\sum_1^n F_k, \overline{C}^*\Big[\sum_1^\infty F_k\Big]\Big) &= D\Big(\sum_1^n C^*[F_k], \overline{C}^*\Big[\sum_1^\infty F_k\Big]\Big) \\ &= D\Big(C^*\Big[\sum_1^n F_k\Big], \overline{C}^*\Big[\sum_1^\infty F_k\Big]\Big) \\ &\leqq WD\Big(\sum_1^n F_k, \sum_1^\infty F_k\Big) < W\epsilon \end{aligned}$$

for all $n \geqq N$. Since the sum $\sum_1^\infty F_k$ is a closed set by agreement, we have $\sum_1^\infty F_k = \overline{C}^*[\sum_1^\infty F_k]$. The proof is complete.

5. **Relations between various types of convergence.** We shall now examine the relations between the various types of convergence defined in §4 and also the relations between the convergence of the two series

$$\sum_1^\infty F_k, \tag{5.1}$$

$$\sum_1^\infty C^*[F_k]. \tag{5.2}$$

(5.3) THEOREM. *The series* (5.2) *converges normally if and only if* (5.1) *converges normally.*

The proof follows from (3.15).

(5.4) THEOREM. *Normal convergence implies strong unconditional convergence.*

The proof follows from Theorem 4.11 and (2.8).

(5.5) THEOREM. *The series* $\sum_1^\infty C^*[F_k]$ *and* $\sum_1^\infty \overline{C}^*[F_k]$ *converge strongly unconditionally if and only if* $\sum_1^\infty F_k$ *converges strongly unconditionally.*

The proof follows from Theorems 4.11, 3.15 and

$$\begin{aligned} D(\overline{C}^*[F_{k(1)}] + \cdots + \overline{C}^*[F_{k(r)}], 0) &= D(C^*[F_{k(1)}] + \cdots + C^*[F_{k(r)}], 0) \\ &= D(C^*[F_{k(1)} + \cdots + F_{k(r)}], 0). \end{aligned}$$

This theorem was given by Birkhoff [1, p. 363].

(5.6) THEOREM. *Strong unconditional convergence implies weak unconditional convergence and regular convergence.*

The proof follows from Theorems 4.11, 4.14 and Definition 4.6.

(5.7) THEOREM. *The series* (5.2) *converges regularly if and only if* (5.1) *converges regularly.*

(5.8) THEOREM. *If a series converges regularly or weakly unconditionally, it converges.*

If a series converges regularly, the sequence of partial sums is a Cauchy sequence.

(5.9) THEOREM. *If* $\sum_1^\infty F_k$ *converges, then* $\sum_1^\infty C^*[F_k]$ *converges and* $\sum_1^\infty C^*[F_k] = \overline{C}^*[\sum_1^\infty F_k]$.

The proof follows from

$$D\Big(\sum_1^n C^*[F_k],\ \overline{C}^*\Big[\sum_1^\infty F_k\Big]\Big) = D\Big(C^*\Big[\sum_1^n F_k\Big],\ \overline{C}^*\Big[\sum_1^\infty F_k\Big]\Big) \leqq WD\Big(\sum_1^n F_k,\ \sum_1^\infty F_k\Big). \tag{5.10}$$

(5.11) THEOREM. *If $\sum_1^\infty F_k$ converges weakly unconditionally, $\sum_1^\infty C^*[F_k]$ converges weakly unconditionally.*

The proof follows from Theorem 4.13 and (5.10).

(5.12) THEOREM. *If $\sum_1^\infty F_k$ is strongly unconditionally convergent with the sum $S\sum_1^\infty F_k$, then it is weakly unconditionally convergent and*

$$\sum_1^\infty F_k = \overline{S\textstyle\sum_1^\infty F_k}.$$

From Theorems 5.6 and 5.8 it follows that $\sum_1^\infty F_k$ converges; let its sum be F. Take any point $f \in F$. We shall show that

$$f \in \overline{S\textstyle\sum_1^\infty F_k}.$$

From $f \in F = \sum_1^\infty F_k$ it follows that corresponding to a given $\epsilon > 0$ there exists an N_1 such that

$$d\Big(f, \sum_1^n F_k\Big) < \epsilon/2, \qquad n \geqq N_1. \tag{5.13}$$

Here $d(g, G)$ denotes the distance from g to G. Since $\sum_1^\infty F_k$ is strongly unconditionally convergent, it follows from Theorem 4.11 that there exists an integer N_2 such that $N_2 < k(1) < \cdots < k(r)$ implies $D(F_{k(1)} + \cdots + F_{k(r)}, 0) < \epsilon/2$, and therefore

$$\Big\| \sum_{i=N_2+1}^\infty f_i \Big\| \leqq \epsilon/2, \qquad f_i \in F_i. \tag{5.14}$$

Let N be the larger of the numbers N_1, N_2. From (5.13) it follows that there exists a sum $\sum_1^N f_k$, $f_k \in F_k$, such that

$$\Big\| f - \sum_1^N f_k \Big\| < \epsilon/2. \tag{5.15}$$

The points $f_1, \cdots, f_N$ are chosen once for all and then held fixed. Then choose points $f_{N+1}, f_{N+2}, \cdots$ arbitrarily from $F_{N+1}, F_{N+2}, \cdots$ and consider the series $\sum_1^\infty f_k$. Since $\sum_1^\infty F_k$ is strongly unconditionally convergent, the series $\sum_1^\infty f_k$ converges and has a sum $g \in S\sum_1^\infty F_k$. But

$$\|f-g\| \leqq \left\| f - \sum_1^N f_k \right\| + \left\| \sum_{N+1}^{\infty} f_k \right\| < \epsilon$$

by (5.15) and (5.14). Since $f \in F, g \in S\sum_1^\infty F_k$ and ϵ is arbitrary, it follows that

$$f \in \overline{S\sum_1^\infty F_k}, \qquad F \subseteqq \overline{S\sum_1^\infty F_k}.$$

We prove next that

$$\overline{S\sum_1^\infty F_k} \subseteqq F.$$

Let $g=\sum_1^\infty f_k, f_k \in F_k$, be an arbitrary point of $S\sum_1^\infty F_k$. Corresponding to any given $\epsilon>0$ we can find an M such that $D(\sum_1^n F_k, F)<\epsilon$ for $n \geqq M$. This fact and the definition of distance imply $d(\sum_1^n f_k, F)<\epsilon$, $n \geqq M$; hence, $d(\sum_1^\infty f_k, F) = d(g, F) \leqq \epsilon$. Since $\epsilon>0$ was arbitrary, $d(g, F)=0$; since F is closed, $g \in F$. Then

$$S\sum_1^\infty F_k \subseteqq F, \qquad \overline{S\sum_1^\infty F_k} \subseteqq F.$$

This inequality completes the proof that

$$\sum_1^\infty F_k = F = \overline{S\sum_1^\infty F_k}$$

and the proof of the theorem.

(5.16) THEOREM. *If the series $\sum_1^\infty F_k$ converges strongly unconditionally, then*

$$C^*\left[S\sum_1^\infty F_k\right] \subseteqq S\sum_1^\infty C^*[F_k] \subseteqq S\sum_1^\infty \overline{C}^*[F_k] \subseteqq \overline{C}^*\left[S\sum_1^\infty F_k\right],$$

$$\overline{S\sum_1^\infty C^*[F_k]} = \overline{S\sum_1^\infty \overline{C}^*[F_k]} = \overline{C}^*\left[S\sum_1^\infty F_k\right].$$

It follows from Theorems 5.5 and 4.15 that $\sum_1^\infty C^*[F_k]$ and $\sum_1^\infty \overline{C}^*[F_k]$ converge strongly unconditionally to sets convex C^*. It is obvious that

$$C^*\left[S\sum_1^\infty F_k\right] \subseteqq S\sum_1^\infty C^*[F_k] \subseteqq S\sum_1^\infty \overline{C}^*[F_k]. \tag{5.17}$$

By Theorem 5.6, $\sum_1^\infty F_k$ converges weakly unconditionally, and by Theorem 5.12 it converges to the sum

$$\overline{S\sum_1^\infty F_k};$$

hence, by Theorems 5.11 and 5.9, $\sum_1^\infty C^*[F_k]$ converges weakly unconditionally to the sum $\overline{C}^*[S\sum_1^\infty F_k]$. Furthermore, $\sum_1^\infty C^*[F_k]$ and $\sum_1^\infty \overline{C}^*[F_k]$ converge

weakly unconditionally to the same sum. Thus $\sum_1^\infty C^*[F_k]$ and $\sum_1^\infty \overline{C}^*[F_k]$ converge strongly unconditionally to $S\sum_1^\infty C^*[F_k]$ and $S\sum_1^\infty \overline{C}^*[F_k]$, and weakly unconditionally to $\overline{C}^*[S\sum_1^\infty F_k]$. It follows from Theorem 5.12 that

$$\overline{S\sum_1^\infty C^*[F_k]} = \overline{S\sum_1^\infty \overline{C}^*[F_k]} = \overline{C}^*\left[S\sum_1^\infty F_k\right].$$

Then $S\sum_1^\infty \overline{C}^*[F_k] \subseteqq \overline{C}^*[S\sum_1^\infty F_k]$. The remainder of the proof follows at once from this fact and (5.17).

This theorem was given by Birkhoff [1, p. 363].

6. **An intersection theorem for series.** The principal object of this section is to establish Theorem 6.4 and two further theorems which give conditions under which its hypothesis are satisfied. This theorem is called an intersection theorem because of its application in the theory of integration. First, we shall state another theorem, whose proof we leave to the reader.

(6.1) THEOREM. *If the series $\sum_1^\infty F_k$ and $\sum_1^\infty G_k$, $G_k \subseteqq F_k$, converge, then $\sum_1^\infty G_k \subseteqq \sum_1^\infty F_k$. If $\sum_1^\infty F_k$ converges regularly, then $\sum_1^\infty G_k$, $G_k \subseteqq F_k$, converges regularly, and $\sum_1^\infty G_k \subseteqq \sum_1^\infty F_k$.*

Let $\{\tau\}$ be a class of transformations τ which satisfies Hypothesis 3.18 and has the additional property that the sum of any unconditionally convergent series of transformations in $\{\tau\}$ is a transformation in $\{\tau\}$. Let C^* be the operator formed from $\{\tau\}$ as explained at the end of §3, and let it satisfy (3.2), (3.3), (3.4).

Consider the infinite series

$$\sum_1^\infty \tau_i F_i, \tag{6.2}$$

$$\sum_{i,j=1}^\infty \tau_{ij} F_{ij}, \tag{6.3}$$

where the F_i and F_{ij} are sets in $\mathfrak{B}$ and the τ_i and τ_{ij} are elements in $\{\tau\}$.

(6.4) THEOREM. *Let the series* (6.2), (6.3) *satisfy the following hypotheses*:

(6.5) $\sum_1^\infty \tau_i F_i$ *converges strongly unconditionally*;

(6.6) $\sum_{j=1}^\infty \tau_{ij}$ *converges unconditionally*, $\sum_{j=1}^\infty \tau_{ij} = \tau_i$ *for all* i;

(6.7) F_i *and* F_{ij} *are convex* C^* *for all* i, j;

(6.8) $F_{ij} \subseteqq F_i$ *for all* i, j; *not all* F_{ij} *are vacuous*; *each* F_i *is bounded.*

Then the following conclusions can be drawn:

(6.9) $\sum_{i,j=1}^\infty \tau_{ij} F_{ij}$ *converges strongly unconditionally*;

(6.10) *each row* $\sum_j \tau_{ij} F_{ij}$ *of* (6 3) *converges strongly unconditionally to a set contained in* $\tau_i F_i$;

(6.11) *the series* $\sum_{ij}\tau_{ij}F_{ij}$ *can be summed by rows*;

(6.12) $\sum_{ij}\tau_{ij}F_{ij} \subseteqq \sum_1^\infty \tau_i F_i$.

Let M be chosen so that $M<k(1)<\cdots<k(r)$ implies $D(\tau_{k(1)}F_{k(1)}+\cdots+\tau_{k(r)}F_{k(r)}, 0)<\epsilon/(2W)$, where W is the constant in (3.5). Such an M exists by the hypotheses of the theorem and Theorem 4.11. Let K denote the maximum of the M finite numbers $D(F_1, 0), \cdots, D(F_M, 0)$. Then choose N so that

$$\left\| \sum \tau_{ij} \right\| < \epsilon/(2MK) \qquad (i = 1, 2, \cdots, M), \tag{6.13}$$

where the summation is with respect to j and over any finite number of terms with $j>N$. Such an N exists by Theorem 4.10 and (6.6).

Now consider any finite sum of terms in $\sum_{ij}\tau_{ij}F_{ij}$ with $i \geqq M+1$ or $j \geqq N+1$. This sum can be broken into two parts, the first of which has $i \leqq M, j \geqq N+1$, and the second of which has $i \geqq M+1$. Consider the first one. We have

$$D\left(\sum_i \sum_j \tau_{ij}F_{ij}, 0\right) \leqq \sum_i D\left(\sum_j \tau_{ij}F_{ij}, 0\right) \leqq \sum_i D\left(\sum_j \tau_{ij}F_i, 0\right)$$

$$\leqq \sum_i D\left(\left(\sum_j \tau_{ij}\right)F_i, 0\right)$$

by (2.8), (6.8), (6.7), and Theorem 3.19. Then from (6.13) we have

$$D\left(\sum_i \sum_j \tau_{ij}F_{ij}, 0\right) \leqq \sum_i \left\| \sum_j \tau_{ij} \right\| D(F_i, 0) < \sum_i \frac{\epsilon}{2MK} K \leqq \epsilon/2.$$

We proceed to show that the second part of the sum is also less than $\epsilon/2$. From (6.8) we have

$$D\left(\sum_i \sum_j \tau_{ij}F_{ij}, 0\right) \leqq D\left(\sum_i \sum_j (\tau_{ij}\tau_i^{-1})\tau_i F_i, 0\right).$$

From (6.6), the hypotheses on $\{\tau\}$ stated above, and the manner of constructing the operator C^*, it follows that $\sum_j(\tau_{ij}\tau_i^{-1})\tau_i F_i \subseteqq C^*[\tau_i F_i\ (+)\ 0]$, where $(+)$ denotes a logical sum rather than a vector sum. Then

$$\begin{aligned} D\left(\sum_i \sum_j \tau_{ij}F_{ij}, 0\right) &\leqq D\left(\sum_i C^*[\tau_i F_i\ (+)\ 0], 0\right) \\ &\leqq D\left(C^*\left[\sum_i (\tau_i F_i\ (+)\ 0)\right], 0\right) \\ &\leqq WD\left(\sum_i (\tau_i F_i\ (+)\ 0), 0\right) \\ &< \epsilon/2 \end{aligned}$$

by Theorems 3.7 and 3.15 and the original choice of M. From (2.8) it then follows that $D(\sum\sum\tau_{ij}F_{ij}, 0)<\epsilon$, where the sum is any finite sum with $i\geqq M+1$ or $j\geqq N+1$. By Theorem 4.11 the series $\sum_{ij}\tau_{ij}F_{ij}$ is strongly unconditionally convergent. The proof of (6.9) is thus complete.

We turn now to (6.10). From Theorem 4.11 and the result just established, it follows that every row $\sum_{ij}\tau_{ij}F_{ij}$ converges strongly unconditionally. Furthermore, $\sum_{j=1}^{J}\tau_{ij}F_{ij} \subseteqq \sum_{j=1}^{J}\tau_{ij}F_i = (\sum_{j=1}^{J}\tau_{ij})F_i$ as a result of hypotheses (6.7), (6.8) and Theorem 3.19. Now by letting $J\to\infty$, we see from (6.6) that the second part of the statement in (6.10) follows.

We proceed to prove (6.11). Let F denote the weak sum of $\sum_{ij}\tau_{ij}F_{ij}$. Since this series converges unconditionally, we can find an m_1 and n_1 so large that

$$D\left(\sum_{i=1}^{r}\sum_{j=1}^{s}\tau_{ij}F_{ij}, F\right) < \epsilon/2 \tag{6.14}$$

for $r\geqq m_1$, $s\geqq n_1$. Next, by Theorem 4.11 and the conclusion (6.9) that $\sum_{ij}\tau_{ij}F_{ij}$ converges strongly unconditionally, there exists an m_2, n_2 such that

$$D(\sum\sum\tau_{ij}F_{ij}, 0) < \epsilon/2, \tag{6.15}$$

where the summation is extended over any set of terms outside of the rectangle $1\leqq i\leqq m_2$, $1\leqq j\leqq n_2$. Let M (N) be the larger of m_1, m_2 (n_1, n_2). Then

$$\begin{aligned} D\left(\sum_{i=1}^{m}\sum_{j=1}^{\infty}\tau_{ij}F_{ij}, F\right) &\leqq D\left(\sum_{i=1}^{m}\sum_{j=1}^{\infty}\tau_{ij}F_{ij}, \sum_{i=1}^{m}\sum_{j=1}^{n}\tau_{ij}F_{ij}\right) \\ &\quad + D\left(\sum_{i=1}^{m}\sum_{j=1}^{n}\tau_{ij}F_{ij}, F\right) \\ &< D\left(\sum_{i=1}^{m}\sum_{j>n}\tau_{ij}F_{ij}, 0\right) + \epsilon/2 \\ &< \epsilon/2 + \epsilon/2 = \epsilon \end{aligned}$$

provided $m\geqq M$, $n\geqq N$ by (2.5), (2.8), (2.4), (6.14), (6.15). This inequality establishes (6.11).

The final conclusion (6.12) now follows from (6.10), (6.11) and the first statement in Theorem 6.1. The proof of the entire theorem is thus complete.

This theorem is an extension of one first given by Birkhoff [1, p. 364].

(6.16) THEOREM. *If the series $\sum_1^\infty\|\tau_i\|$ converges, and if the sets F_i are uniformly bounded, then the series $\sum_1^\infty\tau_iF_i$ converges strongly unconditionally, that is, hypothesis* (6.5) *is satisfied.*

The proof follows from Theorem 4.11.

(6.17) THEOREM. *If $\sum_i\tau_iF_i$ is a finite sum, then hypothesis* (6.5) *is satisfied.*

7. **Summation of infinite series.** We shall now extend some well known theorems on the summation of infinite series to series of the form $\sum_1^\infty f_n$.

Let S denote the sequence $g_1, g_2, \cdots$ of elements in $\mathfrak{B}$. Let (A) denote the following infinite array of numbers:

$$\begin{array}{l} a_{11}, a_{12}, \cdots, a_{1k}, \cdots, \\ a_{21}, a_{22}, \cdots, a_{2k}, \cdots, \\ \cdots\cdots\cdots, \\ a_{i1}, a_{i2}, \cdots, a_{ik}, \cdots, \\ \cdots\cdots\cdots. \end{array}$$

Set $A_i(S)=\sum_{k=1}^\infty a_{ik}g_k$. We say S is summable to the sum $A(S)$ when each of the series $A_i(S)$ is convergent and $A_i(S)\to A(S)$. The method A is called *regular* if every convergent sequence is summable by that method to its limit.

(7.1) THEOREM. *In order that the method A be regular, it is necessary and sufficient that the following conditions be satisfied simultaneously:*

$$\sum_{k=1}^{\infty} |a_{ik}| \leqq M, \qquad i = 1, 2, \cdots; \tag{7.2}$$

$$\lim_{i\to\infty} a_{ik} = 0, \qquad k = 1, 2, \cdots; \tag{7.3}$$

$$\lim_{i\to\infty} \sum_{k=1}^{\infty} a_{ik} = 1. \tag{7.4}$$

The statement of the theorem is identical with the well known theorem for sequences S whose elements are numbers (see Banach [1, pp. 90–91]). The reader will prove without difficulty that the conditions stated are sufficient. By considering sequences of the form $b_1f, b_2f, \cdots$, where f is a fixed element in $\mathfrak{B}$ and $b_1, b_2, \cdots$ is a sequence of numbers, and applying the known theorem for sequences whose elements are numbers, we see that the conditions stated are necessary in the present case also.

(7.5) THEOREM. *If the series $\sum_1^\infty f_n$ is summable C_k, that is, by Cesàro means of order k, and if $f_n=O(1/n)$, then $\sum_1^\infty f_n$ is convergent.*

The proof of this theorem is the same as for a series of numbers (see Knopp [1, pp. 486–487]).

PART III. A THEORY OF INTEGRATION

8. **Abstract space $\mathfrak{A}$.** We shall now consider a space which will be the domain of functions to be studied presently. We assume that $\mathfrak{A}$ is an abstract space with elements x and sets X. In addition we shall suppose that there is a

completely additive class of sets $\mathfrak{X}$ in $\mathfrak{A}$ with the following properties: (1) the empty class belongs to $\mathfrak{X}$; (2) if X belongs to $\mathfrak{X}$, the complement of X does also; (3) the sum of a sequence $\{X_n\}$ of sets selected from the class $\mathfrak{X}$ belongs also to the class $\mathfrak{X}$. We shall say that a set ($\mathfrak{X}$) is measurable ($\mathfrak{X}$). Finally, we shall suppose that there is a measure defined for the class of sets $\mathfrak{X}$. A function of a set $\tau(X)$ will be called a measure ($\mathfrak{X}$) if it has the following properties:

(8.1) $\tau(X)$ is defined for every set in $\mathfrak{X}$; $\tau(X) \in \mathfrak{T}$; if $\tau(X) \neq 0$, then $\tau^{-1}(X)$ exists and $\tau^{-1}(X) \in \mathfrak{T}$.

(8.2) $\tau(\sum_1^\infty X_n) = \sum_1^\infty \tau(X_n)$ for every sequence of sets ($\mathfrak{X}$) no two of which have points in common; the series $\sum_1^\infty \tau(X_n)$ converges unconditionally.

Then $\tau(X)$ is called, for every set X measurable ($\mathfrak{X}$), the measure of X.

We shall not require that the measure of the entire space be finite. We shall admit the case in which the measure $\tau(X)$ of certain sets X in $\mathfrak{X}$ is "infinite" and not a proper transformation in $\mathfrak{T}$. In particular, let $\{X_n\}$ be a denumerable disjoint decomposition of $\mathfrak{A}$ into sets X_n of $\mathfrak{X}$ with measure $\tau(X_n)$. If

$$\lim_{k\to\infty} \left\| \left(\sum_{n=1}^{k} \tau(X_{\alpha(n)}) \right)^{-1} \right\| = 0$$

for every arrangement α of the sets in $\{X_n\}$ and for every such decomposition $\{X_n\}$ of $\mathfrak{A}$, we shall say that the measure of $\mathfrak{A}$ is "infinite," and we shall consider that (8.2) is satisfied.

We shall now describe the construction of the operator C^* to be used in the future. Let $X_1, \cdots, X_r$ be any finite number of disjoint sets in $\mathfrak{X}$ with measures $\tau(X_1), \cdots, \tau(X_r)$ or $\tau_1, \cdots, \tau_r$ for short. Let t, t' be $[(\tau_1 + \cdots + \tau_r)^{-1}\tau_1, \cdots, (\tau_1 + \cdots + \tau_r)^{-1}\tau_r]$, $[\tau_1(\tau_1 + \cdots + \tau_r)^{-1}, \cdots, \tau_r(\tau_1 + \cdots + \tau_r)^{-1}]$ and C_0^* the set of all such t, t'. Then C^* is formed from C_0^* as indicated in §3. We see that C^* automatically satisfies Hypotheses 3.2 and 3.3, and we now make the following further assumption, which is an additional hypothesis concerning the measure function $\tau(X)$:

(8.3) C^* is a bounded operator, that is, C^* satisfies Hypothesis 3.4.

We observe that if the measure function $\tau(X)$ satisfies (8.1), (8.2), (8.3), then the operator C^*, formed as just described, has all the properties assumed in any of the theorems given above. In particular, the assumption was made in §6 that the sum of any unconditionally convergent series of transformations in $\{\tau\}$ is a transformation in $\{\tau\}$; this property is a consequence of (8.2). Hypothesis 3.18 follows from (8.1), (8.2).

Finally, if the measure $\tau(X)$ of a set X in $\mathfrak{X}$ is a positive number $m(X)$, the ordinary convex operator C can be substituted in all later developments

for the operator C^*. In this special case it is clear that for any set F, $C^*[F] \subseteqq C[F]$, and that C has all the desired properties.

We shall associate a second measure function $\nu(X)$ with the class of sets $\mathfrak{X}$. More precisely, let X be any set in $\mathfrak{X}$ and $\{Y_n\}$, $Y_n \in \mathfrak{X}$, any disjoint decomposition of X. Then

$$\nu(X) = \sup_{\{Y_n\}} \sum_{n=1}^{\infty} \|\tau(Y_n)\|. \tag{8.4}$$

If $\tau(X)$ is a positive numerically-valued measure function $m(X)$, then $\nu(X) = \tau(X)$ for every X in $\mathfrak{X}$.

It can be shown that $\nu(X)$ is a measure function with the properties specified above. In particular, $\nu(X)$ has the following properties: (a) $\nu(X)$ is defined for every set $X \in \mathfrak{X}$ and $\nu(X) \geqq 0$; (b) if $\{X_n\}$ is any denumerable set of disjoint sets of $\mathfrak{X}$, then $\nu(\sum_1^\infty X_n) = \sum_1^\infty \nu(X_n)$.

The total variation $\nu(\mathfrak{A})$ of $\tau(X)$ over $\mathfrak{A}$ may be finite. On the other hand it may happen that $\nu(\mathfrak{A})$ is infinite, but that there is a decomposition of $\mathfrak{A}$ into a denumerable set $\{X_n\}$ of disjoint sets of $\mathfrak{X}$ such that $\nu(X_n)$ is finite, $n = 1, 2, \cdots$. These are the only two cases which we shall consider in the future.

We shall say a certain property holds almost everywhere $(\mathfrak{X}, \tau)$ [or $(\mathfrak{X}, \nu)$] and mean thereby that the property in question is satisfied in $\mathfrak{A}$ except at the points of a set $X \in \mathfrak{X}$ with $\tau(X) = 0$ [or $\nu(X) = 0$]. It is clear that almost everywhere $(\mathfrak{X}, \nu)$ implies almost everywhere $(\mathfrak{X}, \tau)$ but not conversely.

9. **Functions.** We shall now describe the class of functions whose integrals we shall consider.

Let $\mathfrak{A}$ be the space described in §8. For each x in $\mathfrak{A}$ let $F(x)$ be a bounded set in $\mathfrak{B}$. Then $F(x)$ is a function whose integral we shall consider in §§10–13 and 20–27. In §§14–19 we shall consider a class of functions $f(x)$ which will be defined as measurable. Here for each x in $\mathfrak{A}$, $f(x)$ is a single element in $\mathfrak{B}$. For the present, however, it is not necessary to restrict ourselves either to measurable functions or to single-valued functions.

10. **Integral ranges.** Let a function $F(x)$ with domain $\mathfrak{A}$ be given. Let Δ_k denote a decomposition of $\mathfrak{A}$ into a denumerably infinite sequence $\{X_i^k\}$ of disjoint sets X_i^k each of which is measurable $(\mathfrak{X})$ and has a measure $\tau(X_i^k)$. The product $\Delta_1\Delta_2$ of two decompositions Δ_1, Δ_2 of $\mathfrak{A}$ is the decomposition of $\mathfrak{A}$ into the sets $X_i^1 X_j^2$ which are nonvacuous.

Consider the series

$$\sum_{i=1}^{\infty} \tau(X_i^k) C^*[F(X_i^k)], \tag{10.1}$$

corresponding to the decomposition Δ_k, where $F(X_i^k)$ is the point-set sum of the sets $F(x)$, $x \in X_i^k$.

(10.2) DEFINITION. *If each set $F(X_i^k)$ is bounded, and if the series* (10.1) *is strongly unconditionally convergent, the sum* (10.1) *will be called the integral range $I(F, \mathfrak{A}, \Delta_k)$ of $F(x)$ with respect to the decomposition Δ_k.*

(10.3) THEOREM. *If the integral ranges $I(F, \mathfrak{A}, \Delta_1)$, $I(F, \mathfrak{A}, \Delta_2)$ exist, then $I(F, \mathfrak{A}, \Delta_1\Delta_2)$ also exists.*

The proof follows from Definition 10.2, Property 8.2, and (6.9) in Theorem 6.4.

(10.4) THEOREM. *Any two integral ranges of $F(x)$ overlap. More precisely, if $I(F, \mathfrak{A}, \Delta_1)$, $I(F, \mathfrak{A}, \Delta_2)$, $I(F, \mathfrak{A}, \Delta_1\Delta_2)$ are integral ranges of $F(x)$ with respect to Δ_1, Δ_2, $\Delta_1\Delta_2$, then*

$$I(F, \mathfrak{A}, \Delta_1\Delta_2) \subseteqq I(F, \mathfrak{A}, \Delta_1)I(F, \mathfrak{A}, \Delta_2).$$

Let the sets in Δ_1, Δ_2 be $\{X_i^1\}$, $\{X_j^2\}$. Then the sets in $\Delta_1\Delta_2$ are $\{X_i^1 X_j^2\}$. Let τ_i and F_i of Theorem 6.4 be identified with $\tau(X_i^1)$ and $C^*[F(X_i^1)]$, and τ_{ij} and F_{ij} with $\tau(X_i^1 X_j^2)$ and $C^*[F(X_i^1 X_j^2)]$. We verify that all the hypotheses of Theorem 6.4 are satisfied. Then from (6.12) we have at once that $I(F, \mathfrak{A}, \Delta_1\Delta_2) \subseteqq I(F, \mathfrak{A}, \Delta_1)$. The remainder of the proof follows by symmetry.

This section follows closely results given by Birkhoff [1, pp. 366–367].

11. **Integrable functions and their integrals.** We proceed now to the definition of integrable function and integral of an integrable function.

(11.1) DEFINITION. *A function $F(x)$ will be called integrable if and only if the inferior limit of the diameters of its integral ranges is zero.*

(11.2) THEOREM. *If $F(x)$ is integrable, then the intersection of the integral ranges of $F(x)$ is a single element $(\mathfrak{X})\int_{\mathfrak{A}} F(x)d\tau(x)$ of $\mathfrak{B}$.*

(11.3) DEFINITION. *The element $(\mathfrak{X})\int_{\mathfrak{A}} F(x)d\tau(x)$ of Theorem* 11.2 *is called the integral of $F(x)$ over $\mathfrak{A}$.*

(11.4) THEOREM. *$F(x)$ is integrable if and only if to every $\epsilon > 0$ there corresponds a decomposition Δ under which the series $\sum_i \tau(X_i)C^*[F(X_i)]$ is strongly unconditionally convergent and its sum has a diameter less than ϵ.*

These definitions and theorems are similar to ones given by Birkhoff [1, p. 367].

12. **Relation to an integral of Birkhoff.** In the special case in which the measure function $\tau(X)$ is a positive numerically-valued function $m(X)$, it is possible to compare the integral $(\mathfrak{X})\int_{\mathfrak{A}} F(x)d\tau(x)$ which we have just defined with an integral $J(F)$ given by Birkhoff [1, p. 367]. In the case now under con-

sideration, therefore, the operator C^* reduces to, or may be replaced by, the ordinary convex operator C.

Birkhoff has defined the integral range $J_\Delta(F)$ of $F(x)$ relative to Δ to be

$$J_\Delta(F) = \overline{C}\left[\mathrm{S}\sum_i m(X_i)F(X_i)\right]. \tag{12.1}$$

His definition of the integral $J(F)$ is identical with that given in Definition 11.3.

(12.2) THEOREM. *If the measure function $\tau(X)$ is a numerically-valued function $m(X)$ and if C^* is replaced by C, then $I(F, \mathfrak{A}, \Delta) = J_\Delta(F)$ and $(\mathfrak{X})\int_{\mathfrak{X}} F(x)dm(x) = J(F)$.*

By Definition 10.2

$$I(F, \mathfrak{A}, \Delta) = \sum_i m(X_i)C[F(X_i)]. \tag{12.3}$$

We observe first from Theorem 5.5 that $\sum_i C[m(X_i)F(X_i)]$, which is the same as (12.3), converges strongly unconditionally if and only if $\sum_i m(X_i)F(X_i)$ converges strongly unconditionally. Hence, the existence of either of the integral ranges implies that of the other.

Next, by Theorem 5.12 we have

$$\sum_i m(X_i)F(X_i) = \overline{\mathrm{S}\sum_i m(X_i)F(X_i)}. \tag{12.4}$$

Again, by Theorem 5.6, we see that the existence of $I(F, \mathfrak{A}, \Delta)$, $J_\Delta(F)$ implies that $\sum_i m(X_i)C[F(X_i)]$ and $\sum_i m(X_i)F(X_i)$ are weakly unconditionally convergent. Finally, from Theorem 5.9 we have

$$\sum_i m(X_i)C[F(X_i)] = \overline{C}\left[\sum_i m(X_i)F(X_i)\right], \tag{12.5}$$

whence, substituting from (12.4) we obtain

$$\begin{aligned}\sum_i m(X_i)C[F(X_i)] &= \overline{C}\left[\mathrm{S}\sum_i m(X_i)F(X_i)\right] \\ &= \overline{C}\left[\mathrm{S}\sum_i m(X_i)F(X_i)\right].\end{aligned} \tag{12.6}$$

Then (12.1), (12.3), (12.6) show that $I(F, \mathfrak{A}, \Delta) = J_\Delta(F)$. From the identity of the integral ranges there follows at once the identity of the integrals. The proof is complete.

It follows from this theorem that the integral defined in §11 genuinely includes that of Birkhoff and reduces to his in a special case.

13. **Properties of the integral.** We shall now give the principal properties of the integral defined in §11.

(13.1) THEOREM. (First theorem on additivity.) *If X is in $\mathfrak{X}$, and if $(\mathfrak{X})\int_{\mathfrak{A}}F(x)d\tau(x)$ exists, then $(\mathfrak{X})\int_{X}F(x)d\tau(x)$ exists. If Δ is a decomposition of $\mathfrak{A}$, and if $(\mathfrak{X})\int_{\mathfrak{A}}F(x)d\tau(x)$ exists, then $\sum_i(\mathfrak{X})\int_{X_i}F(x)d\tau(x)$ converges unconditionally and*

$$\sum_i (\mathfrak{X})\int_{X_i} F(x)d\tau(x) = (\mathfrak{X})\int_{\mathfrak{A}} F(x)d\tau(x). \tag{13.2}$$

(13.3) THEOREM. (Second theorem on additivity.) *If $(\mathfrak{X})\int_{X_i}F(x)d\tau(x)$ exists for all sets X_i of a decomposition Δ of $\mathfrak{A}$, and if $\sum_i(\mathfrak{X})\int_{X_i}F(x)d\tau(x)$ converges unconditionally, then $(\mathfrak{X})\int_{\mathfrak{A}}F(x)d\tau(x)$ exists, and the relation* (13.2) *holds.*

(13.4) THEOREM. (Theorem on distributivity.) *If $F_1(x)$, $F_2(x)$ are integrable over $\mathfrak{A}$ and T_1, T_2 are transformations in $\mathfrak{T}$ which commute with all elements of $\mathfrak{T}$, then $T_1F_1(x)+T_2F_2(x)$ is integrable, and*

$$(\mathfrak{X})\int_{\mathfrak{A}}\left[\sum_{i=1}^{2} T_iF_i(x)\right]d\tau(x) = \sum_{i=1}^{2} T_i\,(\mathfrak{X})\int_{\mathfrak{A}} F_i(x)d\tau(x). \tag{13.5}$$

(13.6) THEOREM. (Theorem on integration of sequences.) *If the functions $F_n(x)$ are defined and integrable over $\mathfrak{A}$, if the $F_n(x)$ tend uniformly to a limit function $F(x)$, and if $\tau(X)$ has bounded total variation $\nu(\mathfrak{A})$ over $\mathfrak{A}$, then $F(x)$ is integrable and*

$$\lim_{n\to\infty} (\mathfrak{X})\int_{\mathfrak{A}} F_n(x)d\tau(x) = (\mathfrak{X})\int_{\mathfrak{A}} F(x)d\tau(x). \tag{13.7}$$

(13.8) THEOREM. (Law of the mean.) *If the measure $\tau(\mathfrak{A})$ of $\mathfrak{A}$ is finite, then*

$$(\mathfrak{X})\int_{\mathfrak{A}} F(x)d\tau(x) = \tau(\mathfrak{A})f, \qquad f \in \overline{C}^*[F(\mathfrak{A})]. \tag{13.9}$$

(13.10) COROLLARY. *If $F(x)$ is bounded in $\mathfrak{A}$ and $\tau(\mathfrak{A})$ is finite, then*

$$\left\| (\mathfrak{X})\int_{\mathfrak{A}} F(x)d\tau(x) \right\| \leqq \|\tau(\mathfrak{A})\| D(C^*[F(\mathfrak{A})], 0). \tag{13.11}$$

The proofs of the first four of the above theorems can be supplied from proofs of similar theorems given by Birkhoff [1, pp. 367–372]. The proof of Theorem 13.8 follows from (10.1) and Theorem 3.19. Each integral range is contained in $\tau(\mathfrak{A})\overline{C}^*[F(\mathfrak{A})]$. Then the intersection of all the integral ranges is a point $\tau(\mathfrak{A})f$ in $\tau(\mathfrak{A})\overline{C}^*[F(\mathfrak{A})]$ and (13.9) follows. Corollary 13.10 follows from (13.9).

(13.12) THEOREM. *Let $F(x)$ be defined in $\mathfrak{A}$, bounded, and equal to the zero element in $\mathfrak{B}$ almost everywhere $(\mathfrak{X}, \tau)$. Then $(\mathfrak{X})\int_{\mathfrak{A}} F(x)d\tau(x)$ exists and equals the zero element.*

PART IV. MEASURABLE FUNCTIONS AND THEIR INTEGRALS

14. **Measurable functions.** We turn now to a consideration of a class of functions to be known as measurable.

(14.1) DEFINITION. *Let $f(x)$ be a single-valued function whose domain is $\mathfrak{A}$ and range is in $\mathfrak{B}$. We shall say that $f(x)$ is measurable $(\mathfrak{X})$ if and only if for every $f_0 \in \mathfrak{B}$ and $r>0$ the set $E_x[\|f(x)-f_0\| \leqq r]$ is measurable $(\mathfrak{X})$.*

We shall consider only functions which, although they may be unbounded, have only finite values.

The definition of measurability in (14.1) can be extended to functions whose ranges are in metric spaces.

(14.2) THEOREM. *If $f(x)$ is measurable, then for all $f_0 \in \mathfrak{B}$ and $r>0$ the following sets are measurable $(\mathfrak{X})$:*

$$E_x[\|f(x) - f_0\| > r], \qquad E_x[\|f(x) - f_0\| = r], \qquad E_x[\|f(x) - f_0\| \geqq r],$$

$$E_x[\|f(x) - f_0\| < r], \qquad E_x[f(x) = f_0].$$

(14.3) THEOREM. *If for all $f_0 \in \mathfrak{B}$ and $r>0$ the set $E_x[\|f(x)-f_0\| <r]$ (or $E_x[\|f(x)-f_0\| \geqq r]$, or $E_x[\|f(x)-f_0\| >r]$) is measurable $(\mathfrak{X})$, then $f(x)$ is measurable $(\mathfrak{X})$.*

We shall have occasion to consider functions of another kind. Let T be a bounded linear transformation with domain $\mathfrak{B}$ and range in $\mathfrak{B}$ and norm $\|T\|$. The space $\mathfrak{T}$ with elements T is a normed vector space. Let $T(x)$ be a function such that $T(x) \in \mathfrak{T}$ for $x \in \mathfrak{A}$. The definition of a function $T(x)$ measurable $(\mathfrak{X})$ can be obtained by replacing $f(x)$ and f_0 in Definition (14.1) by $T(x)$ and T_0. Then f can be replaced by T in Theorems 14.2 and 14.3. A function whose values are real numbers and which is measurable $(\mathfrak{X})$ is a special case both of functions $f(x)$ and $T(x)$ which are measurable $(\mathfrak{X})$.

(14.4) DEFINITION. *Let $f(x)$ be a function with domain $\mathfrak{A}$ and range in $\mathfrak{B}$. We say $f(x)$ is separable if and only if the set $f(\mathfrak{A})$ is separable. The term almost separable $(\mathfrak{X}, \tau)$ [or $(\mathfrak{X}, \nu)$] will have its usual meaning.*

15. **Elementary properties of measurable functions.** In this section we shall establish the fundamental properties of measurable functions.

(15.1) THEOREM. *If $f(x)$ is measurable $(\mathfrak{X})$, then the real-valued function $\|f(x)\|$ is measurable $(\mathfrak{X})$.*

Since $f(x)$ is measurable, the set $E_x[\|f(x)-0\| \leqq r]$ or $E_x[\|f(x)\| \leqq r]$ is measurable.

(15.2) THEOREM. *Let $f(x)$ be measurable $(\mathfrak{X})$. Let N be any positive number, and let a function $f_N(x)$ be defined as follows: $f_N(x)=f(x)$ when $\|f(x)\| \leqq N$ and $f_N(x)=0$ when $\|f(x)\|>N$. Then $f_N(x)$ is measurable $(\mathfrak{X})$.*

Let X_0 be the set on which $f_N(x)=0$. It is measurable $(\mathfrak{X})$ by Theorem 14.2. Then the proof follows from the identities

$$E_x[\|f_N(x)-f_0\| \leqq r] = E_x[\|f(x)-f_0\| \leqq r]\cdot E_x[\|f(x)\| \leqq N] + X_0,$$
$$\text{or } E_x[\|f(x)-f_0\| \leqq r]\cdot E_x[\|f(x)\| \leqq N].$$

The first statement holds in case the sphere $\|f-f_0\| \leqq r$ contains the zero element, the second in case it does not.

(15.3) THEOREM. *If $f_1(x)$ and $f_2(x)$ are measurable $(\mathfrak{X})$, and if one of the functions, say $f_1(x)$, is separable, then $f_1(x)+f_2(x)$ is measurable $(\mathfrak{X})$.*

Let S denote the open sphere $\|f-f_0\|<r$, where $f_0 \in \mathfrak{B}$ and $r>0$ are arbitrary but fixed. To prove that $f_1(x)+f_2(x)$ is measurable, it is sufficient to prove that $E_x[f_1(x)+f_2(x) \in S]$ is measurable (see Theorem 14.3).

From the hypothesis of the theorem it follows that there is a denumerable set of points $g_1, g_2, \cdots$ which is dense in the range of $f_1(x)$. Let $S_i(\xi_k)$ denote the open sphere with center g_i and radius ξ_k, where ξ_k is a rational number such that $0 \leqq \xi_k < r$. Let $S_i'(r-\xi_k)$ denote the open sphere with center f_0-g_i and radius $r-\xi_k$.

The proof of the theorem will be complete when we have established the following identity

$$(15.4)\quad E_x[f_1(x)+f_2(x) \in S] = \sum_{i=1}^{\infty}\sum_{\xi_k} E_x[f_1(x) \in S_i(\xi_k)]\cdot E_x[f_2(x) \in S_i'(r-\xi_k)].$$

The proof of this identity will depend on the following lemma, whose proof we leave to the reader.

(15.5) LEMMA. *The vector sum $S_i(\xi_k)+S_i'(r-\xi_k)$ of the open spheres S_i and S_i' with centers g_i and f_0-g_i and radii ξ_k and $r-\xi_k$ is the open sphere S with center f_0 and radius r.*

From this lemma it is immediately obvious that the set on the left in (15.4) contains the set on the right, and we have to establish only the opposite inequality. Let x_0 be any point of the set on the left in (15.4), that is, a point such that $f_1(x_0)+f_2(x_0) \in S$. We shall show that there is an i and a ξ_k such that

$$x_0 \in E_x[f_1(x) \in S_i(\xi_k)]\cdot E_x[f_2(x) \in S_i'(r-\xi_k)].$$

Since S is open,

$$\|f_1(x_0) + f_2(x_0) - f_0\| = \rho < r. \tag{15.6}$$

Choose ξ_{k_0} as a rational number such that

$$0 < \xi_{k_0} < (r - \rho)/2. \tag{15.7}$$

Next, since the set $g_1, g_2, \cdots$ is dense in the range of $f_1(x)$, there exists a point g_{i_0} such that

$$\|f_1(x_0) - g_{i_0}\| < \xi_{k_0}. \tag{15.8}$$

Thus $S_{i_0}(\xi_{k_0})$ contains $f_1(x_0)$. Then $S'_{i_0}(r-\xi_{k_0})$ contains $f_2(x_0)$, for the contrary assumption gives $\|f_2(x_0)-(f_0-g_{i_0})\| \geqq r-\xi_{k_0}$ and

$$\begin{aligned}\rho = \|f_1(x_0) + f_2(x_0) - f_0\| &\geqq \big|\, \|f_2(x_0) - (f_0 - g_{i_0})\| - \|f_1(x_0) - g_{i_0}\| \,\big| \\ &\geqq \|f_2(x_0) - (f_0 - g_{i_0})\| - \|f_1(x_0) - g_{i_0}\| \\ &> (r - \xi_{k_0}) - \xi_{k_0} = r - 2\xi_{k_0} \\ &> r - (r - \rho) = \rho.\end{aligned}$$

We are able to drop the absolute value signs because $\|f_2(x_0)-(f_0-g_{i_0})\| > r/2$ and $\|f_1(x_0)-g_{i_0}\| < r/2$ by (15.7), (15.8). The assumption has led to a contradiction; hence, $f_2(x_0) \in S'_{i_0}(r-\xi_{k_0})$ as well as $f_1(x_0) \in S_{i_0}(\xi_{k_0})$. Therefore

$$x_0 \in E_x[f_1(x) \in S_{i_0}(\xi_{k_0})] \cdot E_x[f_2(x) \in S'_{i_0}(r - \xi_{k_0})],$$

and any point contained in the set on the left in (15.4) is also contained in the set on the right. This statement completes the proof of the identity (15.4) and the proof of the theorem.

(15.9) THEOREM. *Let $f_1(x), f_2(x), \cdots$ be a sequence of functions measurable $(\mathfrak{X})$ defined on $\mathfrak{A}$ such that $\lim f_n(x)$ exists and equals $f(x)$ for $x \in \mathfrak{A}$. Then $f(x)$ is measurable $(\mathfrak{X})$. If the functions $f_1(x), f_2(x), \cdots$ are separable, then $f(x)$ is separable.*

To prove the theorem we must show that $E_x[\|f(x)-f_0\| \leqq r]$, for arbitrary $f_0 \in \mathfrak{B}$ and $r>0$, is a set $(\mathfrak{X})$. Let $\epsilon_1, \epsilon_2, \cdots$ be a monotone decreasing sequence of positive numbers whose limit is zero. Set

$$X_n(\epsilon) = E_x[\|f_n(x) - f_0\| \leqq r + \epsilon] \cdot E_x[\|f_{n+1}(x) - f_0\| \leqq r + \epsilon] \cdots,$$

$$X(\epsilon) = \sum_1^\infty X_n(\epsilon), \qquad X = \prod_n X(\epsilon_n).$$

It is clear that X is measurable, and that we have to prove only that $E_x[\|f(x)-f_0\| \leqq r] = X$.

Let x_0 be any point of $E_x[\|f(x)-f_0\| \leqq r]$. Then to each ϵ of $\epsilon_1, \epsilon_2, \cdots$ there corresponds an $N=N(f_0, r, x_0, \epsilon)$ such that $\|f_n(x_0)-f_0\| \leqq r+\epsilon$ for $n \geqq N$. Then $x_0 \in X_N(\epsilon)$ and therefore to each $X(\epsilon)$ and finally to X. Conversely, suppose $x_0 \in X$. Then x_0 belongs to each $X(\epsilon)$, and there exists an $N=N(f_0, r, x_0, \epsilon)$ so that $x_0 \in X_N(\epsilon)$. Thus $\|f_n(x_0)-f_0\| \leqq r+\epsilon$ for $n \geqq N$. Then $\|f(x_0)-f_0\| \leqq r+\epsilon$, and since this is true for each ϵ, we have $\|f(x_0)-f_0\| \leqq r$. Therefore $x_0 \in E_x[\|f(x)-f_0\| \leqq r]$.

It is obvious that $f(x)$ is separable under the conditions stated. The proof is complete.

It was pointed out in §14 that the definition of measurable function which we have given applies equally well to functions with values in a metric space. It may be added now that Theorem 15.9 holds also for functions of this kind.

At this point we shall establish the relation between functions which are measurable ($\mathfrak{X}$) and almost separable and two other classes of measurable functions. A function $f(x)$ on $\mathfrak{A}$ to $\mathfrak{B}$ is called a *step-function* if and only if $f(x)$ is constant on each of a finite number of disjoint sets of $\mathfrak{X}$. Then $f(x)$ is measurable (Bochner) if and only if it is almost everywhere the strong limit of a sequence of step-functions (see Bochner [1]). A measurable function which has only a countable number of distinct values will be said to be *countably-valued*. A function $f(x)$ is *weakly measurable* if and only if $u(f(x))$ is measurable for every linear functional $u(f)$ (an additive, continuous, real-valued function) defined in $\mathfrak{B}$ (see Pettis [1, p. 278]). A necessary and sufficient condition that $f(x)$ be measurable (Bochner) is that it be weakly measurable and almost separable (Pettis [1, p. 278]).

(15.10) THEOREM. *A function $f(x)$ is measurable and almost separable if and only if it is the limit almost everywhere of a sequence of step-functions, that is, if and only if it is measurable (Bochner).*

First, a necessary and sufficient condition that an almost separable function $f(x)$ be measurable is that the set $E_x[f(x) \in G]$ be measurable for every open set G in $\mathfrak{B}$. The condition is sufficient by Theorem 14.3, and we shall prove that it is necessary. Without loss of generality, we may assume that $f(x)$ is separable; then $f(\mathfrak{A})G$ contains a denumerable dense set $g_1, g_2, \cdots$. By considering the set of open spheres which are contained in G, have rational radii, and have centers at the points g_i, we see that $E_x[f(x) \in G]$ is the sum of a denumerable number of measurable sets and is therefore measurable. Second, the class of measurable and almost separable functions contains the class of functions which are measurable (Bochner) and is contained in the class of weakly measurable functions. It is clear that a step-function is meas-

urable and separable; then by Theorem 15.9 a function which is measurable (Bochner) is measurable and almost separable. Next, the set $E_f[u(f)<u_0]$, where $u(f)$ is a linear functional defined in $\mathfrak{B}$, is an open set G in $\mathfrak{B}$. Since $E_x[u(f(x))<u_0]=E_x[f(x)\ \varepsilon\ G]$, which is a measurable set by the first step in the proof, it follows that an almost separable, measurable function $f(x)$ is weakly measurable. Third, it is known that for almost separable functions the class of functions which are measurable (Bochner) is the same as the class of functions which are weakly measurable (see Pettis [1, p. 278]). Then the three classes of functions considered in the second step of the proof coincide. The proof is complete.

(15.11) COROLLARY. *An almost separable function is measurable if and only if it is the uniform limit, except possibly on a set of measure zero, of measurable countably-valued functions.*

This corollary follows from the last theorem and a result of Pettis [1, p. 279, Corollary 1.12].

The following corollary is a generalization of the familiar theorem that the product of two measurable functions whose values are numbers is a measurable function.

(15.12) COROLLARY. *If $T(x)$ and $f(x)$ are defined on $\mathfrak{A}$, $T(x)\ \varepsilon\ \mathfrak{T}$, $f(x)\ \varepsilon\ \mathfrak{B}$, are measurable $(\mathfrak{X})$, and have separable ranges, then $T(x)f(x)$ is measurable $(\mathfrak{X})$ and separable.*

(15.13) LEMMA. *If $\mathfrak{A}$ has finite measure $\tau(\mathfrak{A})$ and if $\{f_n(x)\}$ is a sequence of finite measurable functions on $\mathfrak{A}$, converging on this set to a finite measurable function $f(x)$, there exists, for each pair of positive numbers ϵ, η, an integer N and a measurable subset X of $\mathfrak{A}$ such that $\|\tau(X)\|<\eta$ and $\|f_n(x)-f(x)\|<\epsilon$ for every $n>N$ and $x\ \varepsilon\ \mathfrak{A}-X$.*

(15.14) EGOROFF'S THEOREM. *If $\mathfrak{A}$ has finite measure $\tau(\mathfrak{A})$ and if $\{f_n(x)\}$ is a sequence of finite measurable functions defined on $\mathfrak{A}$ that converges almost everywhere $(\mathfrak{X}, \tau)$ on this set to a finite measurable function $f(x)$, then there exists, for each $\epsilon>0$, a subset Y of $\mathfrak{A}$ such that $\|\tau(\mathfrak{A}-Y)\|<\epsilon$, and such that the convergence of $\{f_n(x)\}$ to $f(x)$ is uniform on Y.*

The proofs of Lemma 15.13 and of Egoroff's theorem are similar to those given for functions whose values are numbers (see Saks [1, pp. 17–18]).

16. **Integrals of bounded, measurable, and almost separable functions; an existence theorem.** We shall now prove a theorem which states sufficient conditions for the existence of the integral which was defined in §11. This theorem is a generalization of the familiar one that a function which is bounded and measurable on a set with finite measure is integrable.

(16.1) EXISTENCE THEOREM. *Let $f(x)$ be defined in $\mathfrak{A}$, measurable $(\mathfrak{X})$, almost separable $(\mathfrak{X}, \tau)$, and bounded*:

$$\|f(x)\| < K, \qquad x \in \mathfrak{A}. \tag{16.2}$$

Let the measure function $\tau(X)$ satisfy (8.1), (8.2), (8.3), *and let the measure $\tau(\mathfrak{A})$ of $\mathfrak{A}$ be finite. Finally, let $\tau(X)$ have the following property: if $X_1, \cdots, X_r$ are any r disjoint sets of $\mathfrak{X}$, and if $f_1, \cdots, f_r$ are any r points in $\mathfrak{B}$, there exists a constant W' such that*

$$\left\| \sum_1^r \tau(X_i) f_i \right\| \leqq W' \max_i \|f_i\|. \tag{16.3}$$

Then the integral $(\mathfrak{X}) \int_{\mathfrak{A}} f(x) d\tau(x)$ exists.

Let $X_0 \in \mathfrak{X}$ be the set such that $\tau(X_0) = 0$ and $f(\mathfrak{A} - X_0)$ is separable; let $g_1, g_2, \cdots$ be the denumerable dense set in $f(\mathfrak{A} - X_0)$. Let S_i^k be the closed sphere $\|f - g_i\| \leqq 1/k$. For a fixed k let sets X_i^k be defined as follows:

$$X_1^k = (\mathfrak{A} - X_0) \cdot E_x[f(x) \in S_1^k],$$

$$X_i^k = (\mathfrak{A} - X_0) \cdot E_x[f(x) \in S_i^k] - \sum_{j=1}^{i-1} X_j^k,$$

for $i = 2, 3, \cdots$. Then for a fixed k the sets $X_0, X_1^k, X_2^k, \cdots$ form a decomposition Δ_k of $\mathfrak{A}$, and Δ_k, $k = 1, 2, \cdots$, is a sequence of decompositions of $\mathfrak{A}$.

Corresponding to Δ_k we have the series

$$\sum_{i=1}^{\infty} \tau_i^k C^*[f_i^k], \tag{16.4}$$

where τ_i^k, f_i^k have been written for the longer expressions $\tau(X_i^k), f(X_i^k)$. Since $\tau(\mathfrak{A})$ is finite by hypothesis, and since $\sum_{i=1}^{\infty} \tau_i^k$ converges unconditionally by (8.2), for any $\epsilon > 0$, there exists an M such that $M < i(1) < \cdots < i(r)$ implies $\|\tau_{i(1)}^k + \cdots + \tau_{i(r)}^k\| < \epsilon/(WK)$. Let S denote the sphere $\|f\| \leqq K$. Then by (16.2) and Theorems 3.9, 3.19 we have

$$\begin{aligned} \tau_{i(1)}^k C^*[f_{i(1)}^k] + \cdots + \tau_{i(r)}^k C^*[f_{i(r)}^k] &\leqq \tau_{i(1)}^k C^*[S] + \cdots + \tau_{i(r)}^k C^*[S] \\ &\leqq (\tau_{i(1)}^k + \cdots + \tau_{i(r)}^k) C^*[S]. \end{aligned}$$

From this result and Theorem 3.15 we have

$$\begin{aligned} D(\tau_{i(1)}^k C^*[f_{i(1)}^k] + \cdots + \tau_{i(r)}^k C^*[f_{i(r)}^k], 0) &\leqq \|\tau_{i(1)}^k + \cdots + \tau_{i(r)}^k\| D(C^*[S], 0) \\ &< \epsilon WK/WK = \epsilon. \end{aligned}$$

By Theorem 4.11 the series (16.4) therefore converges strongly uncondition-

ally. Then by Definition 10.2 the sum of (16.4) is an integral range $I(f, \mathfrak{A}, \Delta_k)$.

To complete the proof, it is sufficient, by Theorem 11.4, to show that for each $\epsilon>0$ there exists a decomposition Δ_k of $\mathfrak{A}$ corresponding to which the diameter of the integral range $I(f, \mathfrak{A}, \Delta_k)$ is less than ϵ.

Let I_k denote $I(f, \mathfrak{A}, \Delta_k)$. Then from the definition of the diameter of a set and (2.5), we have

$$\rho(I_k) = D(I_k - I_k, 0) \leqq D\Big(I_k - I_k, \sum_{i=1}^{n} \tau_i^k C^*[f_i^k] - \sum_{i=1}^{n} \tau_i^k C^*[f_i^k]\Big) + D\Big(\sum_{i=1}^{n} \tau_i^k C^*[f_i^k] - \sum_{i=1}^{n} \tau_i^k C^*[f_i^k], 0\Big).$$

Next, by (2.8) and (2.6) we have

$$\rho(I_k) \leqq 2D\Big(I_k, \sum_{i=1}^{n} \tau_i^k C^*[f_i^k]\Big) + D\Big(\sum_{i=1}^{n} \tau_i^k \{C^*[f_i^k] - C^*[f_i^k]\}, 0\Big). \tag{16.5}$$

First choose k so that $k>4WW'/\epsilon$, and let it be held fixed thereafter. Then since f_i^k is contained in the sphere S_i^k with diameter $2/k$, we see that the second term in (16.5) does not exceed

$$W'D(C^*[f_i^k - f_i^k], 0) \leqq 2WW'/k < \epsilon/2. \tag{16.6}$$

But for a fixed k the first term in (16.5) can be made less than $\epsilon/2$ by taking n sufficiently large. We observe that (16.6) is independent of n. We have thus shown that it is possible to choose k so that $\rho(I_k)<\epsilon$. The proof is complete.

(16.7) COROLLARY. *Let $f(x)$ be defined in $\mathfrak{A}$, measurable, almost separable $(\mathfrak{X}, \tau)$ and bounded. Let the measure function $\tau(X)$ satisfy* (8.1), (8.2), (8.3), *and let the measure $\tau(\mathfrak{A})$ of $\mathfrak{A}$ be finite. Finally, let $\tau(X)$ have bounded total variation $\nu(\mathfrak{X})$ over $\mathfrak{A}$, Then the integral $(\mathfrak{X})\int_{\mathfrak{A}} f(x)d\tau(x)$ exists.*

17. **Properties of integrals of bounded, measurable functions.** In §13 we have given a number of properties of the integrals of integrable functions. In this section we shall give certain additional properties possessed by the integrals of bounded measurable functions but not by those of integrable functions in general.

(17.1) THEOREM. (Lebesgue's convergence theorem.) *Let the measure $\tau(\mathfrak{A})$ be finite. Let $f_1(x), f_2(x), \cdots$ be a bounded sequence of functions, defined in $\mathfrak{A}$, which are measurable $(\mathfrak{X})$ and integrable, and which approach almost everywhere $(\mathfrak{X}, \tau)$ an integrable limit function $f(x)$. Then*

$$\lim_{n\to\infty} (\mathfrak{X}) \int_{\mathfrak{A}} f_n(x)d\tau(x) = (\mathfrak{X}) \int_{\mathfrak{A}} f(x)d\tau(x). \tag{17.2}$$

The proof can be given in the usual way by means of Theorems 15.9, 13.4, and 13.1, Lemma 15.13, and Corollary 13.10.

(17.3) COROLLARY. *Let $\tau(\mathfrak{A})$ be finite. Let $f_1(x), f_2(x), \cdots$ be a bounded sequence of functions, defined in $\mathfrak{A}$, which are measurable $(\mathfrak{X})$ and almost separable $(\mathfrak{X}, \tau)$ and approach almost everywhere $(\mathfrak{X}, \tau)$ a limit function $f(x)$. Then $f(x)$ is integrable and the relation* (17.2) *holds.*

From Theorem 15.9 it follows that $f(x)$ is measurable $(\mathfrak{X})$, and it can be shown without difficulty that it is amost separable $(\mathfrak{X}, \tau)$. The remainder of the proof follows from Theorems 16.1 and 17.1.

(17.4) THEOREM. *If $f(x)$ is defined in $\mathfrak{A}$, bounded, measurable $(\mathfrak{X})$, and integrable, and if $\tau(X)$ has bounded total variation $\nu(X)$ over $\mathfrak{A}$, then $(\mathfrak{X})\int_{\mathfrak{A}}\|f(x)\|d\nu(x)$ exists and*

$$\left\| (\mathfrak{X})\int_{\mathfrak{A}} f(x)d\tau(x) \right\| \leqq (\mathfrak{X})\int_{\mathfrak{A}}\|f(x)\|d\nu(x). \tag{17.5}$$

By Theorem 15.1 the function $\|f(x)\|$ is measurable $(\mathfrak{X})$. The existence of $(\mathfrak{X})\int_{\mathfrak{A}}\|f(x)\|d\nu(x)$ then follows from Theorem 16.1. Next, consider the integral range (16.4). We have

$$\begin{aligned} D\left(\sum_{i=1}^{\infty}\tau(X_i^k)C^*[f(X_i^k)], 0\right) &\leqq \sum_{i=1}^{\infty}\|\tau(X_i^k)\|D(C^*[f(X_i^k)], 0) \\ &\leqq \sum_{i=1}^{\infty}\nu(X_i^k)[D(f(X_i^k), 0) + W\rho(f(X_i^k))] \\ &\leqq \sum_{i=1}^{\infty}\nu(X_i^k)\sup_{x\in X_i^k}\|f(x)\| + 2W\nu(\mathfrak{A})/k. \end{aligned}$$

By taking the limit as $k\to\infty$, we obtain (17.5). The proof is complete.

(17.6) COROLLARY. *If $f(x)$ is defined in $\mathfrak{A}$, bounded, measurable $(\mathfrak{X})$, and almost separable $(\mathfrak{X}, \tau)$, and if $(\mathfrak{X})\int_{\mathfrak{A}}\|f(x)\|d\nu(x)$ exists, then* (17.5) *holds.*

There exists a denumerable decomposition $\{X_n\}$ of $\mathfrak{A}$ into sets of $\mathfrak{X}$ such that $\nu(X_n)$ is finite for all n (see §8). Then by the theorem, (17.5) holds for each set X_n. The remainder of the proof follows from Theorem 13.1.

18. **Summable functions and their integrals.** In the preceding two sections we have treated the integrals of bounded, measurable $(\mathfrak{X})$, and almost separable $(\mathfrak{X}, \tau)$ functions. We shall now indicate briefly how these results can be extended to unbounded functions.

(18.1) DEFINITION. *A function $f(x)$, defined in $\mathfrak{A}$, is said to be summable if and only if it is measurable $(\mathfrak{X})$ and almost separable $(\mathfrak{X}, \tau)$ and $(\mathfrak{X})\int_{\mathfrak{A}}\|f(x)\|d\nu(x)$ exists.*

Let $u(x)$ be a real-valued and positive function defined on $\mathfrak{A}$ which is summable (ν). Let $f_1(x), f_2(x), \cdots$ be a sequence of functions defined on $\mathfrak{A}$ each of which is bounded and measurable $(\mathfrak{X})$, almost separable $(\mathfrak{X}, \tau)$ and which approach $f(x)$ almost everywhere $(\mathfrak{X}, \nu)$ and satisfy $\|f_n(x)\| \leqq u(x)$, $x \in \mathfrak{A}$. Then $(\mathfrak{X})\int_{\mathfrak{A}} f_n(x)d\tau(x)$ exists for every n by Theorem 16.1. Furthermore $\lim_{n\to\infty}(\mathfrak{X})\int_{\mathfrak{A}} f_n(x)d\tau(x)$ exists, for we have

$$\left\| (\mathfrak{X})\int_{\mathfrak{A}} f_m(x)d\tau(x) - (\mathfrak{X})\int_{\mathfrak{A}} f_n(x)d\tau(x) \right\| \leqq (\mathfrak{X})\int_{\mathfrak{A}} \|f_m(x) - f_n(x)\| d\nu(x)$$

$$\leqq (\mathfrak{X})\int_{\mathfrak{A}} 2u(x)d\nu(x).$$

There exists a set A, measurable $(\mathfrak{X})$, such that $\mathfrak{A}-A$ has finite measure (ν), and such that $(\mathfrak{X})\int_A 2u(x)d\nu(x) < \epsilon/2$. Also, by Egoroff's theorem and the known properties of summable functions, $(\mathfrak{X})\int_{\mathfrak{A}-A}\|f_m(x) - f_n(x)\| d\nu(x) < \epsilon/2$ for all m, n sufficiently large. Thus for all m, n sufficiently large $\|(\mathfrak{X})\int_{\mathfrak{A}} f_m(x)d\tau(x) - (\mathfrak{X})\int_{\mathfrak{A}} f_n(x)d\tau(x)\| < \epsilon$, and the limit exists as stated. Finally, an argument used by Bochner [1, pp. 266–267] shows that this limit is unique. More precisely, let $v(x)$ and $g_1(x), g_2(x), \cdots$ have the properties specified above for $u(x)$ and $f_1(x), f_2(x), \cdots$. Then

$$\lim_{n\to\infty} (\mathfrak{X})\int_{\mathfrak{A}} g_n(x)d\tau(x) = \lim_{n\to\infty} (\mathfrak{X})\int_{\mathfrak{A}} f_n(x)d\tau(x).$$

Let $f(x)$ be a summable function, and let $f_N(x)$ be the truncated function of $f(x)$ as in Theorem 15.2. Then $f_N(x)$, $N = 1, 2, \cdots$, has all the properties of the sequences just considered. It follows therefore that $\lim_{N\to\infty} (\mathfrak{X})\int_{\mathfrak{A}} f_N(x)d\tau(x)$ exists and is unique.

(18.2) DEFINITION. *Let $f(x)$, $x \in \mathfrak{A}$, be summable, and let $\{f_n(x)\}$, $\|f_n(x)\| \leqq u(x)$, be any sequence of functions, each of which is defined in $\mathfrak{A}$, bounded and measurable $(\mathfrak{X})$, almost separable $(\mathfrak{X}, \tau)$, and which approach $f(x)$ almost everywhere $(\mathfrak{X}, \nu)$. Then the integral of $f(x)$ over $\mathfrak{A}$, denoted by $(\mathfrak{X})\int_{\mathfrak{A}} f(x)d\tau(x)$, is defined by*

$$(18.3) \qquad (\mathfrak{X})\int_{\mathfrak{A}} f(x)d\tau(x) = \lim_{n\to\infty} (\mathfrak{X})\int_{\mathfrak{A}} f_n(x)d\tau(x).$$

The integral of a summable function is effectively defined since the sequence of truncated functions forms one sequence of the needed type.

We shall omit the proofs of the following two theorems.

(18.4) THEOREM. *If $f(x)$ and $g(x)$ are summable, then $f(x)+g(x)$ is summable, and*

$$(\mathfrak{X})\int_{\mathfrak{A}}[f(x)+g(x)]d\tau(x)=(\mathfrak{X})\int_{\mathfrak{A}}f(x)d\tau(x)+(\mathfrak{X})\int_{\mathfrak{A}}g(x)d\tau(x).$$

(18.5) CONVERGENCE THEOREM. *If $f_1(x), f_2(x), \cdots$ is a sequence of summable functions on $\mathfrak{A}$ which approaches almost everywhere $(\mathfrak{X}, \nu)$ a function $f(x)$, and if there exists a summable function $u(x)$ such that $\|f_n(x)\| \leqq u(x)$ for $n=1, 2, \cdots$ and $x \in \mathfrak{A}$, then $f(x)$ is summable and*

$$\lim_{n\to\infty}(\mathfrak{X})\int_{\mathfrak{A}}f_n(x)d\tau(x)=(\mathfrak{X})\int_{\mathfrak{A}}f(x)d\tau(x). \tag{18.6}$$

19. **Comparison with the measurable and summable functions of Bochner and Gowurin.** Other measurable and summable functions have been treated by Bochner [1] and Gowurin [1]. Bochner, however, considered only the case in which the measure function is numerically-valued. In Theorem 15.10 we have shown that a function is measurable and almost separable if and only if it is measurable (Bochner). Furthermore, it follows from the results in §18 that the classes of functions which are summable and summable (Bochner) are one, and that their integrals are the same in the two cases. Other results on the relations of the various integrals are given by Pettis [1, p. 292].

There are functions which are measurable but not measurable (Bochner); one simple example is the everywhere discontinuous, Riemann integrable function given by Graves [1, p. 166] (see Price [2, §4]).

Gowurin's results are patterned after those of Bochner. In his treatment of the Radon integral (see Gowurin [1, pp. 264–265]), which is the part of his paper most closely related to §§14–18, his results are more general than those obtained above in some respects, and less general in others. In particular, his treatment includes only totally measurable functions $f(x)$, where a totally measurable function is bounded and the limit of a uniformly convergent series of functions $g_1(x), g_2(x), \cdots$. On the other hand, some of the restrictions placed on the measure function $\tau(X)$ in §8 are not needed for Gowurin's results. The limit theorems given above (Theorems 17.1, 18.5) are better, disregarding the stronger hypotheses on $\tau(X)$, than that obtained by Gowurin [1, p. 265].

PART V. RIEMANN-STIELTJES INTEGRALS

20. **Definitions and hypotheses.** We proceed to develop the theory of the Riemann-Stieltjes integral. Much of it can be obtained by specializing the general results in Part III. For the sake of simplicity, the exposition will be given for functions defined on a linear interval and a numerically-valued

measure function; from the preceding part of the paper obvious extensions will occur to the reader.

We collect our hypotheses here.

(20.1) HYPOTHESIS. *$F(x)$ is defined on $\mathfrak{A}$: $a \leqq x \leqq b$; $F(x) \subset \mathfrak{B}$.*

(20.2) HYPOTHESIS. *$D(C[F(x)], 0) \leqq K$, $x \in \mathfrak{A}$.*

The measure function $m(I)$ is a numerically-valued, additive, non-decreasing function of intervals $I \subset \mathfrak{A}$:

(20.3) HYPOTHESIS. *$m(I) \geqq 0$, $I \subseteqq \mathfrak{A}$; $0 < m(\mathfrak{A}) < M$, a constant.*

(20.4) HYPOTHESIS. *$m(I_1+I_2) = m(I_1)+m(I_2)$ for any two nonoverlapping intervals I_1, I_2 on $\mathfrak{A}$.*

We denote by I both an interval and its length. Occasionally it will be assumed that $m(I)$ is continuous:

(20.5) HYPOTHESIS. *For every $\epsilon > 0$ there exists an η such that $m(I) < \epsilon$ for every $I \subseteqq \mathfrak{A}$ for which $I < \eta$.*

Since the measure function is numerically-valued, we shall replace the operator C^* by the ordinary convex operator C as explained in §8. Furthermore, Δ_k will now denote a decomposition of $\mathfrak{A}$ into a finite number of nonoverlapping intervals $\{I_i^k\}$, $i = 1, \cdots, n_k$. The integral range of $F(x)$ over $\mathfrak{A}$ with respect to the decomposition Δ_k is

$$I(F, \mathfrak{A}, \Delta_k) = \sum_{i=1}^{n_k}{}^* C[F(I_i^k)]m(I_i^k) = \sum_{i=1}^{n_k}{}^* \overline{C}[F(I_i^k)]m(I_i^k),$$

where $\sum^*$ denotes the closure of the sum (see Definition 10.2). Theorem 6.17 shows that Theorem 10.4 still holds. The definitions and theorems of §11 apply to the present integral $\int_a^b F(x)dm(x)$, which corresponds to one called the generalized Riemann-Stieltjes integral by Hobson [1, vol. 1, p. 547]. We shall now define a set in $\mathfrak{B}$ which generalizes the upper and lower generalized Riemann-Stieltjes integrals (see Hobson [1, vol. 1, p. 546]).

(20.6) DEFINITION. *A set $\mathrm{RS}[\int_a^b F(x)dm(x)]$ such that*

$$D\left(\mathrm{RS}\left[\int_a^b F(x)dm(x)\right],\ I(F, \mathfrak{A}, \Delta_k)\right) \to 0$$

as $k \to \infty$ for every sequence of decompositions Δ_k, $k = 1, 2, \cdots$, for which as $k \to \infty$ $\max_i I_i^k \to 0$ will be called the Riemann-Stieltjes integral set of $F(x)$ on $\mathfrak{A}$.

In the remainder of the paper we propose to study some of the fundamental properties and applications of both $\int_a^b F(x)dm(x)$ and $\mathrm{RS}[\int_a^b F(x)dm(x)]$.

21. **Darboux's theorem and other results.** We shall give first a theorem which generalizes for $\mathrm{RS}[\int_a^b F(x)dm(x)]$ the well known theorems of Darboux for upper and lower Riemann integrals (see Hobson [1, vol. 1, pp. 462–463]).

As a matter of notation, let $\Delta_k' : \{I_i'^k\}$ $(i=1, \cdots, n_k)$, $k=1, 2, \cdots$, denote a sequence of consecutive decompositions of $\mathfrak{A}$, that is, a sequence in which Δ_{k+1}' can be obtained from Δ_k' by subdividing one or more intervals. It will be assumed for all sequences Δ_k, Δ_k' that $\max_i I_i^k, \max_i I_i'^k \to 0$ as $k \to \infty$.

(21.1) THEOREM. *Let $F(x)$ and $m(I)$ satisfy Hypotheses* 20.1–20.5. *If there exists a set $G \subset \mathfrak{B}$ such that $D(G, I(F, \mathfrak{A}, \Delta_k')) \to 0$ as $k \to \infty$, then $D(G, I(F, \mathfrak{A}, \Delta_k)) \to 0$ as $k \to \infty$ for every sequence of decompositions $\Delta_1, \Delta_2, \cdots$, and G is the Riemann-Stieltjes integral set* $\mathrm{RS}[\int_a^b F(x)dm(x)]$.

Although not difficult, the details of the proof are long and will be omitted.

(21.2) THEOREM. *Let $F(x)$ and $m(I)$ satisfy Hypotheses* 20.1–20.5. *If $F(\mathfrak{A})$ is compact in $\mathfrak{B}$, then* $\mathrm{RS}[\int_a^b F(x)dm(x)]$ *exists.*

To prove the theorem we shall show that there exists a set G such that $D(G, I(F, \mathfrak{A}, \Delta_k')) \to 0$ as $k \to \infty$. It will then follow from Theorem 21.1 that $\mathrm{RS}[\int_a^b F(x)dm(x)]$ exists and is identical with G.

From Theorem 10.4 and the fact that the decompositions are consecutive,

$$I(F, \mathfrak{A}, \Delta_1') \supseteq I(F, \mathfrak{A}, \Delta_2') \supseteq \cdots . \tag{21.3}$$

Since $F(\mathfrak{A})$ is compact in $\mathfrak{B}$ by hypothesis, it follows that $C[F(\mathfrak{A})]$ is compact (see Mazur [1] or Price [1, §7]). From this fact it follows that each set in (21.3) is compact. Under these conditions Cantor's theorem states that the product G of the sets in (21.3) is not null (see Sierpiński [1, p. 30] or Hausdorff [1, p. 129]). Finally, since each set of (21.3) is closed and convex, G is closed and convex.

We shall show next that $D(G, I(F, \mathfrak{A}, \Delta_k')) \to 0$ as $k \to \infty$. Suppose that this limit is not zero. Since $G \subseteq I(F, \mathfrak{A}, \Delta_k')$, $k=1, 2, \cdots$, the distance from any point of G to $I(F, \mathfrak{A}, \Delta_k')$ is zero. Then for some $\epsilon > 0$ there exist sets $G_k \subset I(F, \mathfrak{A}, \Delta_k')$, $k=1, 2, \cdots$, such that the distance from each point of G_k to G is ϵ or greater, for otherwise the limit would be zero. For each k the set G_k is closed, compact, and non-null, and $G_1 \supseteq G_2 \supseteq \cdots$. Then by Cantor's theorem the product G' of the sets G_k is non-null. Finally, the distance from any point of G' to G is at least ϵ. But since $G' \subseteq G$, this is impossible, and therefore $D(G, I(F, \mathfrak{A}, \Delta_k')) \to 0$ as $k \to \infty$. The proof is complete.

(21.4) COROLLARY. *The Riemann-Stieltjes integral set is closed and convex.*

(21.5) THEOREM. *Let $F(x)$ and $m(I)$ satisfy Hypotheses* 20.1–20.4. *A necessary and sufficient condition that* $\mathrm{RS}[\int_a^b F(x)dm(x)]$ *exist is that for every se-*

quence of subdivisions Δ_k, $k=1, 2, \cdots$, *of* $\mathfrak{A}$ *the sets* $I(F, \mathfrak{A}, \Delta_k)$ *form a Cauchy sequence.*

A necessary condition that a sequence converge is that it be a Cauchy sequence; the necessity of the condition stated in the theorem then follows from Definition 20.6. Also the condition is sufficient. Consider the two sequences of subdivisions Δ_k, Δ_k^*. Since the space whose elements are sets in $\mathfrak{B}$ is complete in the Hausdorff metric (see §2), each of the Cauchy sequences $I(F, \mathfrak{A}, \Delta_k)$, $I(F, \mathfrak{A}, \Delta_k^*)$ has a limit. If these limits are not the same, the sequence of decompositions $\Delta_1, \Delta_1^*, \Delta_2, \Delta_2^*, \cdots$ leads to a sequence of integral ranges which is not a Cauchy sequence. This fact contradicts the hypothesis. The proof is complete.

22. **Existence theorems.** We shall now give two theorems which state sufficient conditions of another type for the existence of $\int_a^b F(x)dm(x)$ and $\mathrm{RS}\left[\int_a^b F(x)dm(x)\right]$.

First, we shall give a definition of continuity. Let I be a closed interval on $\mathfrak{A}$. Then the fluctuation $\Phi(F, I)$ of $F(x)$ on I is the diameter $\rho(F(I)) = \rho(C[F(I)])$ of $F(I)$. In terms of the fluctuation we can define the saltus $\omega(F, x_0)$ at a point $x_0 \in \mathfrak{A}$. Let I_k be a closed interval of length $2k$ whose midpoint is x_0, the proper modification being made in case x_0 is an end point of $\mathfrak{A}$. The limit of $\Phi(F, I_k)$ as $k \to 0$ exists and is unique, and is $\omega(F, x_0)$ by definition. We shall say that $F(x)$ is continuous at x_0 if and only if $\omega(F, x_0)=0$. Thus if $F(x)$ is continuous at x_0, $F(x_0)$ denotes a single element in $\mathfrak{B}$.

With this definition of continuity, we have the following theorem.

(22.1) THEOREM. *Let* $F(x)$ *and* $m(I)$ *satisfy Hypotheses* 20.1–20.4. *If the variation of* $m(I)$ *over the points of discontinuity of* $F(x)$ *is zero, then the Riemann-Stieltjes integral* $\int_a^b F(x)dm(x)$ *exists.*

The proof of this theorem can be patterned very closely after a well known existence theorem for the Riemann-Stieltjes integral of a numerically-valued function (see Hobson [1, pp. 542–544]). In the present case the proof is somewhat simpler than that given by Hobson as a result of the general theorems proved above (see §§10, 11). We show that $\rho(I(F, \mathfrak{A}, \Delta_k)) \to 0$ as $k \to \infty$ for an arbitrary sequence Δ_k, $k=1, 2, \cdots$. It follows that $\int_a^b F(x)dm(x)$ exists, and that $I(F, \mathfrak{A}, \Delta_k) \to \int_a^b F(x)dm(x)$ for an arbitrary sequence Δ_k.

The corresponding theorem for the Riemann integral in abstract spaces was first proved by Graves [1].

We shall now give a second definition of continuity and from it obtain another existence theorem. The two definitions of continuity are distinct, and they lead to different existence theorems.

The fluctuation $\Phi(F, I)$ of $F(x)$ on the closed interval I is defined by

$$\Phi(F, I) = \sup_{x_1, x_2 \in I} D(C[F(x_1)], C[F(x_2)]). \tag{22.2}$$

The saltus $\omega(F, x_0)$ of $F(x)$ at $x_0 \in \mathfrak{A}$ is defined in terms of this fluctuation in the usual way. Finally, $F(x)$ is continuous at $x_0 \in \mathfrak{A}$ if and only if $\omega(F, x_0) = 0$. A point of continuity of $F(x)$ according to this definition may be a point of discontinuity according to the first definition.

(22.3) THEOREM. *Let $F(x)$ and $m(I)$ satisfy Hypotheses* 20.1–20.4. *If the variation of $m(I)$ over the points of discontinuity (in the second sense) of $F(x)$ is zero, then the Riemann-Stieltjes integral set* $\mathrm{RS}[\int_a^b F(x)dm(x)]$ *exists.*

As a result of Theorem 21.5, this theorem will be established if we show that $I(F, \mathfrak{A}, \Delta_k)$, $k = 1, 2, \cdots$, is a Cauchy sequence for every sequence of decompositions Δ_k. The analysis used in establishing this fact is similar to that in the proof of Theorem 22.1 and is omitted.

The conditions stated in Theorems 22.1, 22.3 are sufficient but are not necessary (see Graves [1]).

23. **Properties of the Riemann-Stieltjes integral set.** We shall first investigate the additive properties of $\mathrm{RS}[\int_a^b F(x)dm(x)]$ with respect to intervals and functions.

It will be convenient to use $F +^* G$ to denote the closure of $F+G$.

(23.1) THEOREM. *If* $\mathrm{RS}[\int_a^c F(x)dm(x)]$ *and* $\mathrm{RS}[\int_c^b F(x)dm(x)]$ *exist, $a < c < b$, then* $\mathrm{RS}[\int_a^b F(x)dm(x)]$ *exists and* $\mathrm{RS}[\int_a^b] = \mathrm{RS}[\int_a^c] + \mathrm{RS}[\int_c^b]$.

Let $\mathfrak{A}_1$, $\mathfrak{A}_2$ denote the intervals $a \leqq x \leqq c$, $c \leqq x \leqq b$. By (2.7), (2.8) we have

$$D\left(\mathrm{RS}\left[\int_a^c F(x)dm(x)\right] +^* \mathrm{RS}\left[\int_a^b F(x)dm(x)\right],\ I(F, \mathfrak{A}_1, \Delta_k') +^* I(F, \mathfrak{A}_2, \Delta_k'')\right) \leqq D\left(\mathrm{RS}\left[\int_a^c\right],\ I(F, \mathfrak{A}_1, \Delta_k')\right) + D\left(\mathrm{RS}\left[\int_c^b\right],\ I(F, \mathfrak{A}_2, \Delta_k'')\right),$$

from which the proof follows.

(23.2) THEOREM. *If integrals* $\mathrm{RS}[\int_a^b F_1(x)dm(x)]$, $\mathrm{RS}[\int_a^b F_2(x)dm(x)]$, *and* $\mathrm{RS}[\int_a^b \{F_1(x) + F_2(x)\}dm(x)]$ *exist, then*

$$\mathrm{RS}\left[\int_a^b \{F_1(x) + F_2(x)\}dm(x)\right] \leqq \mathrm{RS}\left[\int_a^b F_1(x)dm(x)\right] +^* \mathrm{RS}\left[\int_a^b F_2(x)dm(x)\right].$$

Let $G(x)=F_1(x)+F_2(x)$. Then $G(I)\subseteqq F_1(I)+F_2(I)$ for any interval $I\subseteqq\mathfrak{A}$. The proof follows from this fact and the properties of the convex operator C.

(23.3) THEOREM. *Let $u(x)$, $v(x)$ be numerically-valued functions such that $u(x)\geqq 0$, $v(x)\geqq 0$. For each $x\in\mathfrak{A}$ let $F(x)$ be a convex set in $\mathfrak{B}$. If* $\mathrm{RS}[\int_a^b u(x)F(x)dm(x)]$ *and* $\mathrm{RS}[\int_a^b v(x)F(x)dm(x)]$ *exist, then* $\mathrm{RS}[\int_a^b\{u(x)+v(x)\}F(x)dm(x)]$ *also exists, and*

$$\mathrm{RS}\left[\int_a^b\{u(x)+v(x)\}F(x)dm(x)\right] = \mathrm{RS}\left[\int_a^b\{u(x)F(x)+v(x)F(x)\}dm(x)\right]. \tag{23.4}$$

(23.5) THEOREM. *Let $u(x)$, $v(x)$ be functions defined on $\mathfrak{A}$ whose values are numbers but otherwise arbitrary. For each $x\in\mathfrak{A}$ let $F(x)$ be an element in $\mathfrak{B}$. If* $\mathrm{RS}[\int_a^b u(x)F(x)dm(x)]$, $\mathrm{RS}[\int_a^b v(x)F(x)dm(x)]$ *exist, then* $\mathrm{RS}[\int_a^b\{u(x)+v(x)\}F(x)dm(x)]$ *exists and* (23.4) *holds.*

The proofs of these two theorems are omitted.

We shall now examine analogues of removing a constant from under the sign of integration.

(23.6) THEOREM. *Let T be a continuous linear transformation with domain and range $\mathfrak{B}$ which has an inverse T^{-1} with the same properties. If* $\mathrm{RS}[\int_a^b F(x)dm(x)]$ *exists, then* $\mathrm{RS}[\int_a^b TF(x)dm(x)]$ *exists and equals* $T\{\mathrm{RS}[\int_a^b F(x)dm(x)]\}$.

The corresponding theorem for the Lebesgue integral was proved byBirkhoff [1, pp. 370–371].

(23.7) THEOREM. *Let $u(x)\geqq 0$ be a bounded, numerically-valued function defined on $\mathfrak{A}$, and F an arbitrary bounded set in $\mathfrak{B}$. Let $m(I)$ satisfy Hypotheses 20.3, 20.4. If* $\mathrm{RS}[\int_a^b u(x)dm(x)]$ *exists, then* $\mathrm{RS}[\int_a^b u(x)Fdm(x)]$ *exists and equals* $\mathrm{RS}[\int_a^b u(x)dm(x)]\overline{C}[F]$.

From the definition of the integral range in §20 and Theorems 2.18, 2.21 we have

$$\begin{aligned} I(u(x)F,\mathfrak{A},\Delta_k) &= \sum_i{}^*\overline{C}[u(I_i^k)F]m(I_i^k) = \sum_i{}^*\overline{C}[u(I_i^k)]\overline{C}[F]m(I_i^k)\\ &= \overline{C}\left\{\sum_i{}^*\overline{C}[u(I_i^k)]m(I_i^k)\right\}\overline{C}[F]\\ &= \left\{\sum_i{}^*\overline{C}[u(I_i)]m(I_i^k)\right\}\overline{C}[F]\\ &= I(u(x),\mathfrak{A},\Delta_k)\overline{C}[F]. \end{aligned}$$

Next, by Theorem 2.21 we have

$$D\left(\mathrm{RS}\left[\int_a^b u(x)dm(x)\right]\overline{C}[F],\ I(u(x)F,\ \mathfrak{A},\ \Delta_k)\right)$$
$$\leqq D\left(\mathrm{RS}\left[\int_a^b u(x)dm(x)\right],\ I(u(x),\ \mathfrak{A},\ \Delta_k)\right)D(\overline{C}[F],\ 0).$$

Since $I(u(x), \mathfrak{A}, \Delta_k)\to\mathrm{RS}[\int_a^b u(x)dm(x)]$ as $k\to\infty$ and F is bounded, the proof follows.

(23.8) THEOREM. *Let $u(x)$ be a bounded, numerically-valued function defined on $\mathfrak{A}$; let $m(I)$ satisfy Hypotheses* 20.3, 20.4; *and let f be an arbitrary element of $\mathfrak{B}$. If $\mathrm{RS}[\int_a^b u(x)dm(x)]$ exists, then $\mathrm{RS}[\int_a^b u(x)fdm(x)]$ exists and equals $\mathrm{RS}[\int_a^b u(x)dm(x)]f$.*

The proof follows from Corollary 2.17.

We shall now prove the analogue of the inequality $|\int_a^b u(x)dx-\int_a^b v(x)dx|\leqq\int_a^b|u(x)-v(x)|dx$ for numerically-valued functions $u(x)$, $v(x)$.

(23.9) THEOREM. *Let $F_1(x)$, $F_2(x)$ and $m(I)$ satisfy Hypotheses* 20.1–20.4. *If $\mathrm{RS}[\int_a^b F_i(x)dm(x)]$, $i=1, 2$, exist, then*

$$D\left(\mathrm{RS}\left[\int_a^b F_1(x)dm(x)\right],\ \mathrm{RS}\left[\int_a^b F_2(x)dm(x)\right]\right)$$
$$\leqq\overline{\int_a^b} D(C[F_1(x)],\ C[F_2(x)])dm(x). \tag{23.10}$$

The integral on the right in (23.10) is the upper generalized Riemann-Stieltjes integral of a numerically-valued function (see Hobson [1, p. 546]).

To prove the theorem, we observe first that

$$D\left(\mathrm{RS}\left[\int_a^b F_1(x)dm(x)\right],\ \mathrm{RS}\left[\int_a^b F_2(x)dm(x)\right]\right)$$
$$\leqq D\left(\mathrm{RS}\left[\int_a^b F_1(x)dm(x)\right],\ I(F_1,\ \mathfrak{A},\ \Delta_k)\right)$$
$$+ D(I(F_1,\ \mathfrak{A},\ \Delta_k),\ I(F_2,\ \mathfrak{A},\ \Delta_k))$$
$$+ D(I(F_2,\ \mathfrak{A},\ \Delta_k),\ \mathrm{RS}\left[\int_a^b F_2(x)dm(x)\right]).$$

Then by (2.6), (2.7), (2.8), and the definition of $I(F, \mathfrak{A}, \Delta_k)$ we have

$$D(I(F_1,\ \mathfrak{A},\ \Delta_k),\ I(F_2,\ \mathfrak{A},\ \Delta_k))\leqq\sum_i D(C[F_1(I_i^k)],\ C[F_2(I_i^k)])m(I_i^k).$$

Finally, we observe that

$$D(C[F_1(I_i^k)],\ C[F_2(I_i^k)]) \leqq \sup_{x \in I_i^k} D(C[F_1(x)],\ C[F_2(x)]).$$

From these facts the proof follows.

The next theorem states a sufficient condition that the Riemann-Stieltjes integral set be additive (see Theorem 23.2). In it and in its proof, we shall understand fluctuation, saltus, and continuity in the second of the two senses of §22 (see (22.2)).

(23.11) THEOREM. *Let $F_1(x)$, $F_2(x)$, and $m(I)$ satisfy Hypotheses* 20.1–20.4, *and let the variation of $m(I)$ over the points of discontinuity of $F_1(x)$ and $F_2(x)$ be zero. Then*

$$\text{RS}\left[\int_a^b \{F_1(x) + F_2(x)\}\, dm(x)\right] = \text{RS}\left[\int_a^b F_1(x) dm(x)\right] +^* \text{RS}\left[\int_a^b F_2(x) dm(x)\right], \tag{23.12}$$

the existence of these integrals following from Theorem 22.3.

The proof of this theorem depends on the following two lemmas, whose proof we leave to the reader.

(23.13) LEMMA. *If $\Phi(F, I)$ is defined by* (22.2), *then $\Phi(F_1+F_2, I) \leqq \Phi(F_1, I) + \Phi(F_2, I)$.*

(23.14) LEMMA. *Let $F_1(x)$, $F_2(x)$ be constant, that is, let $F_1(x) = F_1$, $F_2(x) = F_2$ for all $x \in \mathfrak{A}$. Then the relation* (23.12) *holds.*

We proceed to the proof of the theorem. Let ζ and ϵ be any positive numbers. Then with each point ξ of $\mathfrak{A}$ we can associate an interval I of length not greater than ζ such that $\Phi(F_i, I) \leqq \omega(F_i, \xi) + \epsilon/2$, $i = 1, 2$ (see Hobson [1, pp. 466–467]). Then by Lemma 23.13, $\Phi(F_1+F_2, I) \leqq \omega(F_1, \xi) + \omega(F_2, \xi) + \epsilon$. By the Heine-Borel theorem a finite number of these intervals cover $\mathfrak{A}$. Finally, we can replace these overlapping intervals by a finite set of non-overlapping intervals I such that in each of them $\Phi(F_1+F_2, I) \leqq \omega(F_1, \xi) + \omega(F_2, \xi) + \epsilon$, where ξ is either an interior or end point of I (see Hobson [1, pp. 466–467]). Let $\eta \leqq \zeta$ be the length of the shortest of these intervals, and let their end points be $a = x_0 < x_1 < \cdots < x_n = b$.

Let two functions $G_1(x)$, $G_2(x)$ be defined on $\mathfrak{A}$ as follows: $G_i(x) = F_i(x_{k-1})$, $x_{k-1} \leqq x < x_k$, for $k = 1, 2, \cdots, n$ and $i = 1, 2$. Set $G(x) = G_1(x) + G_2(x)$.

Let $A(\epsilon)$ be the closed set of points of $\mathfrak{A}$ at which the saltus of either $F_1(x)$ or $F_2(x)$ is greater than or equal to ϵ. With each point $x \in A(\epsilon)$ associate the in-

terval $x-\rho \leqq x \leqq x+\rho$. These intervals coalesce to form a finite number of nonoverlapping intervals contained in $\mathfrak{A}$; denote them by $A(\epsilon, \rho)$. The set $A(\epsilon)$ is the inner limiting set of the sets $A(\epsilon, \rho)$ as $\rho \to 0$. Since the variation of $m(I)$ over $A(\epsilon)$ is zero by hypothesis, it is possible to choose ρ so that the variation of $m(I)$ over $A(\epsilon, \rho)$ is less than an arbitrary positive number ζ'. Choose ζ so that $\zeta < \rho$; then also $\eta < \rho$. Then if one of the intervals I contains a point of $A(\epsilon)$, it is contained entirely in $A(\epsilon, \rho)$. Let $\mathfrak{A}_1$ denote the sum of the intervals I which contain no point of $A(\epsilon)$ and $\mathfrak{A}_2$ the sum of the remaining ones. Then $m(\mathfrak{A}_2) < \zeta'$.

We can now establish (23.12) by showing that the distance between the sets on the two sides of the equation is zero. By (2.5) we have

$$
\begin{aligned}
&D\Big(\mathrm{RS}\Big[\int_a^b \{F_1(x)+F_2(x)\}dm(x)\Big],\ \mathrm{RS}\Big[\int_a^b F_1(x)dm(x)\Big] \\
&\qquad +^*\mathrm{RS}\Big[\int_a^b F_2(x)dm(x)\Big]\Big) \\
&\leqq D\Big(\mathrm{RS}\Big[\int_a^b \{F_1(x)+F_2(x)\}dm(x)\Big],\ \mathrm{RS}\Big[\int_a^b G(x)dm(x)\Big]\Big) \\
&\qquad + D\Big(\mathrm{RS}\Big[\int_a^b G(x)dm(x)\Big],\ \mathrm{RS}\Big[\int_a^b F_1(x)dm(x)\Big] \\
&\qquad\qquad +^*\mathrm{RS}\Big[\int_a^b F_2(x)dm(x)\Big]\Big).
\end{aligned} \tag{23.15}
$$

By Theorems 23.1, 22.3, and (2.8) the first term in the right-hand member of (23.15) does not exceed

$$
\begin{aligned}
&D\Big(\mathrm{RS}\Big[\int_{\mathfrak{A}_1} \{F_1(x)+F_2(x)\}dm(x)\Big],\ \mathrm{RS}\Big[\int_{\mathfrak{A}_1} G(x)dm(x)\Big]\Big) \\
&\qquad + D\Big(\mathrm{RS}\Big[\int_{\mathfrak{A}_2} \{F_1(x)+F_2(x)\}dm(x)\Big],\ DS\Big[\int_{\mathfrak{A}_2} G(x)dm(x)\Big]\Big).
\end{aligned} \tag{23.16}
$$

By Theorem 23.9 and the results above, the first term here does not exceed

$$
\begin{aligned}
\int_{\mathfrak{A}_1} D(C[F_1(x)+F_2(x)], C[G(x)])dm(x) &\leqq \int_{\mathfrak{A}_1} [\omega(F_1, \xi)+\omega(F_2, \xi)+\epsilon]dm(x) \\
&\leqq 3\epsilon m(\mathfrak{A}_1) \leqq 3\epsilon m(\mathfrak{A}).
\end{aligned}
$$

The upper integral of Theorem 23.9 is an ordinary integral here by Theorem 22.1 and the hypotheses of the present theorem. Also by Theorem 23.9, (2.5), Hypothesis 20.2, and the results above, the second term in (23.16) does not exceed

$$\int_{\mathfrak{A}_2} D(C[F_1(x)+F_2(x)], C[G(x)])dm(x) \leqq \int_{\mathfrak{A}_2} 4Kdm(x) < 4K\zeta'.$$

Thus the first term in the right-hand member of (23.15) is less than $3\epsilon m(\mathfrak{A})+4K\zeta'$.

Now consider the second term in the right member of (23.15). By Lemma 23.14 and the definition of $G(x)$ we have $\mathrm{RS}[\int_a^b G(x)dm(x)] = \mathrm{RS}[\int_a^b G_1(x)dm(x)] + {}^*\mathrm{RS}[\int_a^b G_2(x)dm(x)]$. Then by (2.7) and (2.8), the second term in the right-hand member of (23.15) does not exceed

$$D\left(\mathrm{RS}\left[\int_a^b G_1(x)dm(x)\right], \mathrm{RS}\left[\int_a^b F_1(x)dm(x)\right]\right) + D\left(\mathrm{RS}\left[\int_a^b G_2(x)dm(x)\right], \mathrm{RS}\left[\int_a^b F_2(x)dm(x)\right]\right);$$

as before we can show that this expression does not exceed $3\epsilon m(\mathfrak{A})+4K\zeta'$.

Collecting results, we see that the left-hand member of (23.15) is less than $6\epsilon m(\mathfrak{A})+8K\zeta'$; since $m(\mathfrak{A})$, K are constants and ϵ, ζ' are arbitrary, it is therefore zero. The proof is complete.

We turn next to a theorem of a different type. Let T denote an element of $\mathfrak{T}$ (see §3). We shall consider a transformation $T(x)$ which depends on the real parameter x. Then $\|[T(x_1)-T(x_2)]f\| \leqq C(x_1, x_2)\|f\|$. We shall say that $T(x)$ is continuous in x at x_0 if and only if for each $\epsilon>0$ there exists a $\delta(x_0, \epsilon)$ such that $C(x, x_0)<\epsilon$ for $|x-x_0|<\delta(x_0, \epsilon)$.

(23.17) THEOREM. *Let $F(x)$ and $m(I)$ satisfy Hypotheses* 20.1–20.4, *and let $T(x)$ be continuous in x on the closed interval $\mathfrak{A}$. If $\mathrm{RS}[\int_c^d F(x)dm(x)]$ exists for every interval $c \leqq x \leqq d$ on $\mathfrak{A}$, then $\mathrm{RS}[\int_a^b T(x)F(x)dm(x)]$ also exists.*

First we approximate to $T(x)$ by a sequence $T_n(x)$, $n=1, 2, \cdots$, such that $T_n(x)$ is constant on subintervals of $\mathfrak{A}$. Then the proof is completed by means of Theorem 23.1, a slight extension of Theorem 23.6, and Theorem 13.6.

24. **Improper integrals.** Let $F(x)$ and $m(I)$ satisfy Hypotheses 20.1, 20.3, 20.4, and let $F(x)$ be bounded except at $x=c$, $a<c<b$. If $\lim \mathrm{RS}[\int_a^{c-\epsilon_1} F(x)dm(x)]$, $\lim \mathrm{RS}[\int_{c+\epsilon_2}^b F(x)dm(x)]$ exist as $\epsilon_1, \epsilon_2 \to 0$ independently, then we say that the improper integral $\mathrm{RS}[\int_a^b F(x)dm(x)]$ exists and that its value is the closure of the sum of these limits. If

$$\lim \overline{\int_a^{c-\epsilon_1}} D(C[F(x)], 0)dm(x), \quad \lim \overline{\int_{c+\epsilon_2}^b} D(C[F(x)], 0)dm(x)$$

exist as $\epsilon_1, \epsilon_2 \to 0$ independently, we shall say that the improper integral $\mathrm{RS}[\int_a^b F(x)dm(x)]$ converges absolutely. Similar considerations apply if there are any finite number of points of $\mathfrak{A}$ at which $F(x)$ is unbounded.

(24.1) THEOREM. *Let $T(x)$ satisfy the hypotheses of Theorem* 23.17. *If $\mathrm{RS}[\int_a^b F(x)dm(x)]$ is an improper integral which converges absolutely, then $\mathrm{RS}[\int_a^b T(x)F(x)dm(x)]$ exists and converges absolutely.*

25. **Extreme integrals.** We shall show that there are certain points of the Riemann-Stieltjes set $\mathrm{RS}[\int_a^b F(x)dm(x)]$ which are exact analogues of the upper and lower Riemann-Stieltjes integrals of a numerically-valued function.

Let $F(x)$ be a function defined and bounded on $\mathfrak{A}$ which has its values in n-dimensional euclidean space but is otherwise arbitrary. By Theorem 2.14 and §20, we have $I(F, \mathfrak{A}, \Delta_k) = \sum_i \overline{C}[F(I_i^k)]m(I_i^k)$. Assume that $\mathrm{RS}[\int_a^b F(x)dm(x)]$ exists. Then the closed convex set $\mathrm{RS}[\int_a^b F(x)dm(x)]$ has at least two extreme points (unless it is a single point), and they are limit points of extreme points of $I(F, \mathfrak{A}, \Delta_k)$ (see Price [1, §9]). Furthermore, if g is an extreme point of $I(F, \mathfrak{A}, \Delta_k)$, then $g = \sum_i g_i m(I_i^k)$, and by Theorem 2.15 g_i is an extreme point of $\overline{C}[F(I_i^k)]$, $i = 1, \cdots, n_k$. Moreover, the extreme points g_i are contained in the closure of the sets $F(I_i^k)$. Finally, the set $\mathrm{RS}[\int_a^b F(x)dm(x)]$ is the closed convex hull of its extreme points.

For a numerically-valued function, the set

$$I(u, \mathfrak{A}, \Delta_k) = \sum_i \overline{C}[u(I_i^k)]m(I_i^k) \tag{25.1}$$

has exactly two extreme points, and they are

$$\sum_i \left[\sup_{x \in I_i^k} u(x)\right] m(I_i^k), \qquad \sum_i \left[\inf_{x \in I_i^k} u(x)\right] m(I_i^k). \tag{25.2}$$

Also, $\sup_{x \in I_i^k} u(x)$ and $\inf_{x \in I_i^k} u(x)$ are extreme points of the set $\overline{C}[u(I_i^k)]$ and belong to the closure of $u(I_i^k)$. Moreover, the limits of the two extreme points (25.2) of (25.1) are

$$\overline{\int_a^b} u(x)dm(x), \qquad \underline{\int_a^b} u(x)dm(x),$$

the two extreme integrals of $\mathrm{RS}[\int^b u(x)dm(x)]$. Finally, the set $\mathrm{RS}[\int_a^b u(x)dm(x)]$ is the interval bounded by—that is, the closed convex hull of—the points

$$\overline{\int_a^b} u(x)dm(x), \qquad \underline{\int_a^b} u(x)dm(x).$$

These results hold at least in n-dimensional euclidean space and possibly

in more general spaces (see Price [1, §9] and Theorems 2.14, 2.15 above).

It was in connection with the results of this paragraph that the author rediscovered extreme points of convex sets.

Part VI. Other measurable functions

26. **An extension of Bochner's results.** We shall now show how the program of Bochner can be carried through in a more general case.

Let $F(x)$, $G(x)$, $\cdots$ denote functions defined for x in $\mathfrak{A}$ whose values are sets in $\mathfrak{B}$. Let $m(X)$ be a numerically-valued, nonnegative measure function defined for sets X in $\mathfrak{X}$. Let $G(x)$ denote a function which takes on only a finite number of values, that is, a function for which there exists a finite decomposition of $\mathfrak{A}$ into sets $X_1, \cdots, X_r$ of $\mathfrak{X}$ with $G(x)=G_i \subset \mathfrak{B}$ for $x \in X_i$, $i=1, \cdots, r$. A function $F(x)$ will be called measurable if there exists a sequence $G_n(x)$ of these functions such that $\lim_{n\to\infty} D(F(x), G_n(x))=0$ almost everywhere $(\mathfrak{X}, m)$.

(26.1) Definition. $\int_{\mathfrak{A}} G(x)dm(x)=\sum_i^* C[G_i]m(X_i)$.

With this definition, we see that the integral of $G_1(x)+G_2(x)$ exists, and that $\int_{\mathfrak{A}}\{G_1(x)+G_2(x)\}dm(x)=\int_{\mathfrak{A}}G_1(x)dm(x)+^*\int_{\mathfrak{A}}G_2(x)dm(x)$. The real-valued function $D(G(x), 0)$ is measurable. From (2.5) we have $|D(F(x), 0) - D(G_n(x), 0)| \leqq D(F(x), G_n(x))$, from which it follows that $D(F(x), 0)$ also is measurable. We see then that $\int_{\mathfrak{A}} D(G(x), 0)dm(x)$ exists, and from (2.8) it follows that $D(\int_{\mathfrak{A}} G(x)dm(x), 0) \leqq \int_{\mathfrak{A}} D(G(x), 0)dm(x)$.

We shall say that the measurable function $F(x)$ is summable if $D(F(x), 0)$ is summable. Let $u(x)$ be a positive, real-valued, summable function; and let $\{F_n(x)\}$ be a sequence of bounded, measurable functions such that $D(F_n(x), 0) \leqq u(x)$ for all n. By arguments similar to those in §18, we can show that if $F(x)$ is summable and $F_n(x) \to F(x)$ almost everywhere $(\mathfrak{X}, m)$, then $\lim \int_{\mathfrak{A}} F_n(x)dm(x)$ exists and is unique. We define $\int_{\mathfrak{A}} F(x)dm(x)$ to be the closure of $\lim \int_{\mathfrak{A}} F_n(x)dm(x)$; it is therefore a closed convex set. Summable functions and their integrals have the following properties:

(26.2) $D(\int_{\mathfrak{A}} F(x)dm(x), 0) \leqq \int_{\mathfrak{A}} D(F(x), 0)dm(x)$;

(26.3) if $F(x)$ is constant and equal to F, then $\int_{\mathfrak{A}} F(x)dm(x)=\overline{C}[F]m(\mathfrak{A})$;

(26.4) if c_1, c_2 are constants and $F_1(x)$, $F_2(x)$ are summable functions, then

$$\int_{\mathfrak{A}} \{c_1F_1(x) + c_2F_2(x)\}dm(x) = c_1\int_{\mathfrak{A}} F_1(x)dm(x) +^* c_2\int_{\mathfrak{A}} F_2(x)dm(x);$$

(26.5) if $\mathfrak{A}=X_1+X_2+\cdots$ and $X_mX_n=0$ for $m\neq n$, then

$$\int_{\mathfrak{A}} F(x)dm(x) = \sum_1^\infty \int_{X_n} F(x)dm(x);$$

(26.6) if the summable functions $F_1(x), F_2(x), \cdots$ on $\mathfrak{A}$ have a summable real-valued bound $u(x)$, that is, if $D(F_n(x), 0) \leqq u(x)$ for all n, and if $\lim F_n(x)$ exists for almost all x, then $\lim \int_{\mathfrak{A}} F_n(x)dm(x) = \int_{\mathfrak{A}} \lim F_n(x)dm(x)$.

27. **A more general class of measurable functions.** Bochner [1] and Dunford [1] have obtained extensive classes of measurable functions from classes of finitely-valued functions and continuous functions respectively. It is possible to show by means of examples that a more general class of measurable functions can be obtained by starting from a more general class. We shall describe such a class of functions.

Let $G(x), G_1(x), \cdots$ be functions defined on $\mathfrak{A}$: $a \leqq x \leqq b$ whose values are sets in $\mathfrak{B}$. Let Φ be a class of these functions with the following properties:

(27.1) $G(x) \in \Phi$ implies $D(G(x), 0)$ is bounded and continuous except possibly on a set of measure zero.

(27.2) $G(x) \in \Phi$ implies $\mathrm{RS}[\int^b G(x)dx]$ exists;

(27.3) $G_1(x), G_2(x) \in \Phi$ implies $G_1(x)+G_2(x) \in \Phi$.

(27.4) $G_1(x), G_2(x) \in \Phi$ implies the points of discontinuity of $D(G_1(x), G_2(x))$ are a set of measure zero at most.

(27.5) $G_1(x), G_2(x) \in \Phi$ implies $\mathrm{RS}[\int_a^b \{G_1(x)+G_2(x)\}dx] = \mathrm{RS}[\int_a^b G_1(x)dx] +^* \mathrm{RS}[\int_a^b G_2(x)dx]$.

The extension of the class Φ and the determination of the properties of the integrals of functions in the extended class are left to the reader.

Part VII. Fourier series

28. **Fejér's theorem.** For functions $f(x)$ which are measurable and almost separable, the fundamental results for singular integrals hold as for numerically-valued functions. The proofs are so similar that they need not be repeated here (see Hobson [1, vol. 2, chap. 7] and Bochner [1, §7]). We shall give certain results for the Fejér integral of functions $F(x)$, however.

Let $F(x)$ be defined for $-\pi \leqq x < \pi$ and elsewhere by $F(x+2\pi) = F(x)$; the values of $F(x)$ are sets in $\mathfrak{B}$. Let $\mathrm{RS}[\int_I F(t)dt]$ exist for every interval I contained in $-\pi \leqq t \leqq \pi$, and if it is an improper integral, let it be absolutely convergent. The Fejér integral corresponding to this function is

$$\frac{1}{2\pi m} \mathrm{RS}\left[\int_{-\pi}^{\pi} \frac{\sin^2 [m(x-t)/2]}{\sin^2 [(x-t)/2]} F(t)dt\right]. \tag{28.1}$$

The existence of this integral, not only for the interval indicated but also for every interval contained in it, for all m and x follows from the hypotheses concerning $F(x)$ and Theorem 24.1.

(28.2) Theorem. *The limit as $m \to \infty$ of the Fejér integral* (28.1) *exists for*

each value of x for which $C[F(x\pm 0)]$ exist, and the value of this limit is $(1/2)\{C[F(x+0)]+{}^*C[F(x-0)]\}$.

The proof of this theorem follows the general outline of the proof of the corresponding theorem for numerically-valued functions. In the course of the proof it is necessary to use (2.5), (2.6), (2.7), the second definition of continuity in §22, and Theorems 23.1, 23.7, 23.9. The details are somewhat long and will be omitted.

Let $f(x)$ satisfy all the hypotheses imposed on $F(x)$ above; in addition let $f(x)$ for each x be an element in $\mathfrak{B}$, and let its points of discontinuity form a set of measure zero. Then $\int_{-\pi}^{\pi} f(t)dt$ exists, and Theorem 28.2 holds for $f(x)$. Let elements a_n, b_n in $\mathfrak{B}$ be defined by

$$(28.3)\qquad a_n = \frac{1}{\pi}\int_{-\pi}^{\pi} f(t)\cos nt\,dt, \qquad b_n = \frac{1}{\pi}\int_{-\pi}^{\pi} f(t)\sin nt\,dt$$

for $n=0, 1, 2, \cdots$. These integrals exist by Theorem 24.1. With these coefficients we form the Fourier series

$$(28.4)\qquad a_0/2 + \sum_{n=1}^{\infty}(a_n\cos nx + b_n\sin nx).$$

The investigation of the summation by Cesàro means of order one (see Knopp [1, chap. 13]) of (28.4) leads in the usual way to Fejér's integral (28.1). Theorems 23.5, 23.6, and 23.11 enable us to make the necessary transformations. Then from Theorem 28.2 we have the following theorem.

(28.5) THEOREM. *Let $f(x)$ have the properties just specified. Then the Fourier series* (28.4) *associated with $f(x)$ is summable C_1 at all points x at which the two limits $f(x\pm 0)$ exist, and its sum C_1 is* $(1/2)[f(x+0)+f(x-0)]$.

Bochner [1, §7] has obtained a more general result by using the Lebesgue integral rather than the Riemann integral. In the present case a theorem more general than (28.5) might be expected, in particular, a theorem for functions $F(x)$ rather than $f(x)$, the series (28.4) having sets rather than elements as coefficients, corresponding to the greater generality in Theorem 28.2. The more general result is lacking because of the restrictive hypotheses in Theorems 23.3, 23.5, and 23.11. It seems therefore that there may be functions which can be represented by the limits of their Fejér integrals but which cannot be represented by their Fourier series.

29. **A theorem on the Fourier coefficients.** Bochner [1, §7] has shown that the coefficients a_n, b_n in (28.3) tend to zero as $n\to\infty$ for any function which is measurable and almost separable. The following generalization of a

well known theorem has a certain amount of interest because it is necessary to construct a new proof.

(29.1) THEOREM. *Let $f(x)$ be a function defined on $\mathfrak{A}$ whose values are elements in $\mathfrak{B}$; let $f(x)$ have bounded total variation on $\mathfrak{A}$. Then as $\lambda \to \infty$*

$$\int_a^b f(x) \sin \lambda x \, dx, \qquad \int_a^b f(x) \cos \lambda x \, dx \tag{29.2}$$

are $O(1/\lambda)$.

Since $f(x)$ has bounded total variation on $\mathfrak{A}$, its points of discontinuity are denumerable and of the first kind only; hence the integrals in (29.2) exist by Theorem 22.1. We shall set up a sequence of subdivisions of the interval $\mathfrak{A}$ which can be used in constructing the integral $\int_a^b f(x) \sin \lambda x \, dx$. On the x axis mark the points $2\pi m/(2^k\lambda)$, $m = 0, \pm 1, \pm 2, \cdots$. The intervals or parts of intervals of this subdivision of the x axis which lie on $\mathfrak{A}$ will be taken as the intervals of the decomposition Δ_k of $\mathfrak{A}$, $k = 1, 2, \cdots$. The decompositions of $\mathfrak{A}$ have been chosen in such a way that the zeros of $\sin \lambda x$ on $\mathfrak{A}$ coincide with end points of intervals. Let the end points of intervals of Δ_k taken from left to right be x_i^k, $i = 0, 1, \cdots, n_k$, and let Δx_i^k denote the length of the interval $x_{i-1}^k \leqq x \leqq x_i^k$.

It is clear that

$$\begin{aligned} \left\| \int_a^b f(x) \sin \lambda x \, dx \right\| \leqq & \left\| \sum_i f(x_i^k) \sin \lambda x_i^k \, \Delta x_i^k \right\| \\ & + \left\| \sum_i f(x_i^k) \sin \lambda x_i^k \, \Delta x_i^k - \int_a^b f(x) \sin \lambda x \, dx \right\|. \end{aligned} \tag{29.3}$$

Since $\int_a^b f(x) \sin \lambda x \, dx$ exists, for each $\epsilon > 0$ it is possible to find a k_0 such that

$$\left\| \sum_i f(x_i^k) \sin \lambda x_i^k \Delta x_i^k - \int_a^b f(x) \sin \lambda x \, dx \right\| < \epsilon \tag{29.4}$$

for $k \geqq k_0$ (see Definition 11.3).

Next consider the first term in the right-hand member of (29.3). Set $s_i = \sin \lambda x_1^k \, \Delta x_1^k + \cdots + \sin \lambda x_i^k \, \Delta x_i^k$ for $i = 1, 2, \cdots, n_k$. Then

$$\begin{aligned} \sum_i f(x_i^k) \sin \lambda x_i^k \Delta x_i^k &= f(x_1^k) s_1 + \sum_{i=2}^{n_k} f(x_i^k)(s_i - s_{i-1}) \\ &= s_1[f(x_1^k) - f(x_2^k)] + s_2[f(x_2^k) - f(x_3^k)] \\ &\quad + \cdots + s_{n_k} f(x_{n_k}^k), \end{aligned}$$

from which it follows that

$$\left\| \sum_i f(x_i^k) \sin \lambda x_i^k \Delta x_i^k \right\| \leqq |s_1| \, \|f(x_1^k) - f(x_2^k)\| + |s_2| \, \|f(x_2^k) - f(x_3^k)\| + \cdots + |s_{n_k}| \, \|f(x_{n_k}^k)\|.$$

From the definition of s_i it follows that $|s_i| < \pi/\lambda$ for $i = 1, 2, \cdots, n_k$. Then since $f(x)$ has bounded total variation on $\mathfrak{A}$, there is a constant V such that

$$\left\| \sum_i f(x_i^k) \sin \lambda x_i^k \Delta x_i^k \right\| < \pi V/\lambda. \tag{29.5}$$

Collecting results from relations (29.3), (29.4), and (29.5) we have $\|\int_a^b f(x) \sin \lambda x \, dx\| < \pi V/\lambda + \epsilon$. Since ϵ is an arbitrary positive number, it follows that $\|\int_a^b f(x) \sin \lambda x \, dx\| \leqq \pi V/\lambda$. These results and similar considerations for the other integral in (29.2) complete the proof.

30. **Fourier's theorem.** From Theorems 7.5, 28.5, and 29.1 we have the following theorem.

(30.1) FOURIER'S THEOREM. *Let $f(x)$ be defined on $-\pi \leqq x < \pi$ and elsewhere by $f(x+2\pi) = f(x)$; let the values of $f(x)$ be elements in $\mathfrak{B}$. If $f(x)$ has bounded total variation on $-\pi \leqq x \leqq \pi$, then the Fourier series* (28.4) *associated with $f(x)$ converges for each value of x to* $(1/2)[f(x+0)+f(x-0)]$.

BIBLIOGRAPHY

STEFAN BANACH

1. *Théorie des Opérations Linéaires*, Monografje Matematyczne, vol. 1, Warsaw, 1932.

GARRETT BIRKHOFF

1. *Integration of functions with values in a Banach space*, these Transactions, vol. 38 (1935), pp. 357–378.
2. *Moore-Smith convergence in general topology*, Annals of Mathematics, (2), vol. 38 (1937), pp. 39–56.

W. BLASCHKE

1. *Kreis und Kugel*, Leipzig, Veit, 1916.

S. BOCHNER

1. *Integration von Funktionen, deren Werte die Elemente eines Vektorraumes sind*, Fundamenta Mathematicae, vol. 20 (1933), pp. 262–276.

NELSON DUNFORD

1. *Integration in general analysis*, these Transactions, vol. 37 (1935), pp. 441–453.
2. *Integration of vector-valued functions*, Bulletin of the American Mathematical Society, vol. 43 (1937), p. 24.
3. *Integration and linear operations*, these Transactions, vol. 40 (1936), pp. 474–494.

MARK GOWURIN

1. *Über die Stieltjessche Integration abstrakter Funktionen*, Fundamenta Mathematicae, vol. 27 (1936), pp. 255–268.

LAWRENCE M. GRAVES

1. *Riemann integration and Taylor's theorem in general analysis*, these Transactions, vol. 29 (1927), pp. 163–177.

F. HAUSDORFF

1. *Mengenlehre*, 2d edition, Berlin and Leipzig, de Gruyter, 1927.

E. W. Hobson
1. *The Theory of Functions of a Real Variable*, vol. 1, 3d edition, 1927; vol. 2, 2d edition, 1926; Cambridge University Press.

S. Mazur
1. *Über die kleinste konvexe Menge, die eine gegebene kompakte Menge enthält*, Studia Mathematica, vol. 2 (1930), pp. 7–9.

Konrad Knopp
1. *Theorie und Andwendung der unendlichen Reihen*, 2d edition, Berlin, Springer, 1924.

B. J. Pettis
1. *On integration in vector spaces*, these Transactions, vol. 44 (1938), pp. 277–304.

G. Baley Price
1. *On the extreme points of convex sets*, Duke Mathematical Journal, vol. 3 (1937), pp. 56–67.
2. *Definitions and properties of monotone functions*, Bulletin of the American Mathematical Society, vol. 46 (1940), pp. 75-78.
3. *On the completeness of a certain metric space with an application to Blaschke's selection theorem*, to appear in the Bulletin of the American Mathematical Society.

Stanislaw Saks
1. *Theory of the Integral*, Monografje Matematyczne, vol. 7, Warsaw and Lwów, 1937.

Waclaw Sierpiński
1. *Introduction to General Topology*, Toronto, The University of Toronto Press, 1934.

Brown University,
Providence, R. I.,
University of Kansas,
Lawrence, Kans.

ON THE GRADIENT OF SOLID HARMONIC POLYNOMIALS*

BY

G. SZEGÖ

To Leopold Fejér
on his 60th birthday, February 9, 1940.

INTRODUCTION

In 1912 S. Bernstein proved a theorem on trigonometric polynomials which has been considered by various authors during subsequent years.† For our later purposes it is convenient to formulate this theorem in terms of harmonic polynomials as follows:

I. *Let $u(x, y)$ be a real harmonic polynomial of the nth degree which satisfies the inequality $|u(x, y)| \leqq 1$ in the unit circle $x^2+y^2 \leqq 1$. If l denotes the positive tangential direction at an arbitrary point (x, y) of the unit circle $x^2+y^2=1$, we have*

$$\left| \frac{\partial u}{\partial l} \right| \leqq n, \qquad x^2 + y^2 = 1. \tag{1}$$

The equality sign holds only if $u(x, y) = \cos n(\phi-\phi_0)$ for $x^2+y^2=1$, where $(1, \phi)$ are the polar coordinates of the point with the cartesian coordinates (x, y) on the unit circle, and ϕ_0 is arbitrary real.

Ten years ago I proved the following theorem, which is more informative than I:‡

II. *Let $u(x, y)$ be a real harmonic polynomial of the nth degree which satisfies the inequality $|u(x, y)| \leqq 1$ in the unit circle $x^2+y^2 \leqq 1$. Then*

$$|\operatorname{grad} u| = (u_x^2 + u_y^2)^{1/2} \leqq n, \qquad x^2 + y^2 \leqq 1. \tag{2}$$

The equality sign holds for the same polynomials as in I; *it holds only for $x^2+y^2=1$ if $n>1$.*

In the present investigation I am dealing with the three-dimensional analogue of the last theorem. The following result holds:

* Presented to the Society, September 8, 1939; received by the editors September 6, 1939.

† See the literature quoted in the introduction of the paper referred to in the next footnote.

‡ G. Szegö, *Über einen Satz des Herrn Bernstein*, Schriften der Königsberger Gelehrten Gesellschaft, 1928, pp. 59–70. See also S. Bernstein, *Sur un théorème de M. Szegö*, Prace Matematyczno-Fizyczne, vol. 44 (1937), pp. 9–14.

BOSTON UNIVERSITY
COLLEGE OF LIBERAL ARTS
LIBRARY

III. Let $u(x, y, z)$ *be a real harmonic polynomial of the nth degree which satisfies the inequality* $|u(x, y, z)| \leqq 1$ *in the unit sphere* $x^2+y^2+z^2 \leqq 1$. *We assume* $n \geqq 4$. *Then*

$$|\operatorname{grad} u| = (u_x^2 + u_y^2 + u_z^2)^{1/2}$$

$$\leqq 2n\left(\frac{1}{1} - \frac{1}{3} + \cdots + \frac{(-1)^{n-1}}{2n-1}\right) + \begin{cases} 0 & \text{if } n \text{ even}, \\ -1 & \text{if } n \text{ odd}, \end{cases} \tag{3}$$

$$x^2 + y^2 + z^2 \leqq 1.$$

The equality sign holds only for the polynomial $u(x, y, z)$ *whose boundary values on the unit sphere* $x^2+y^2+z^2=1$ *are of the form* $\pm\cos n\gamma$; *here* γ *is the spherical distance of the variable point* (x, y, z) *on the unit sphere from a fixed point* P_0 *of this sphere.*

For these polynomials the equality sign holds only at P_0 *and at the point diametral to* P_0.

The bound ρ_n in the inequality (3) is the precise one. We have $\rho_n \cong n\pi/2$ as $n \to \infty$.

The proof follows an argument very much similar to that used in the proof of my former Theorem II. The underlying idea is due essentially to M. and F. Riesz; these authors dealt, of course, only with the inequality (1) of S. Bernstein. In the present case the chief difficulty is to prove the positivity of the trigonometric polynomial

$$M_n(\theta) = 1 + 2\sum_{\nu=1}^{n-1}\left\{\frac{\rho_{n-\nu}}{\rho_n}\cos n\theta \cos(n-\nu)\theta + [(n-\nu)/n]\sin n\theta \sin(n-\nu)\theta\right\} \tag{4}$$

for all real values of θ.

Unfortunately, my proof for (3) fails if $n=2$ or $n=3$. In these cases the less precise inequality

$$|\operatorname{grad} u| \leqq 2^{1/2}n \tag{3'}$$

can be obtained. We notice that $2^{1/2}\cdot 2 = 2.82\cdots$ whereas $\rho_2 = 8/3 = 2.66\cdots$, and $2^{1/2}\cdot 3 = 4.23\cdots$ whereas $\rho_3 = 21/5 = 4.20$.

First part of the proof

1. It is sufficient to prove (3) for $x=y=0$, $z=1$. For the other points of the unit sphere the statement follows by a simple rotation, for the interior points by considering $u(kx, ky, kz)$, $0 \leqq k < 1$.

If we introduce polar coordinates r, θ, ϕ (in the usual notation), we have for $u(x, y, z) = U(r, \theta, \phi)$

$$|\operatorname{grad} u| = \{U_r^2 + (r^{-1}U_\theta)^2 + ((r\sin\theta)^{-1}U_\phi)^2\}^{1/2}, \tag{5}$$

so that our inequality (3) is equivalent to the following statement: Let α, β be arbitrary real constants, and

$$K = \cos\alpha \cdot U_r + \sin\alpha\cos\beta \cdot r^{-1}U_\theta + \sin\alpha\sin\beta(r\sin\theta)^{-1}U_\phi; \tag{6}$$

then

$$|K| \leqq \rho_n. \tag{7}$$

2. Representing $u(x, y, z) = U(r, \theta, \phi)$ in the form

$$U(r, \theta, \phi) = A_0(r, \theta) + \sum_{\nu=1}^{n} (A_\nu(r, \theta)\cos\nu\phi + B_\nu(r, \theta)\sin\nu\phi), \tag{8}$$

we have

$$\begin{aligned} A_\nu(r, \theta) &= \sin^\nu\theta \sum_{\mu=\nu}^{n} a_{\mu\nu} P_\mu^{(\nu)}(\cos\theta) r^\mu, \\ B_\nu(r, \theta) &= \sin^\nu\theta \sum_{\mu=\nu}^{n} b_{\mu\nu} P_\mu^{(\nu)}(\cos\theta) r^\mu, \end{aligned} \tag{9}$$

where $a_{\mu\nu}$ and $b_{\mu\nu}$ are real constants; $P_\mu(\cos\theta)$ is Legendre's polynomial in the usual notation.

For $x = y = 0$, $z = 1$, or $r = 1$, $\theta = 0$, we find

$$\begin{aligned} U_r &= \frac{\partial A_0}{\partial r} = \sum_{\mu=1}^{n} \mu a_{\mu 0}, \\ r^{-1}U_\theta &= \frac{\partial A_1}{\partial \theta}\cos\phi + \frac{\partial B_1}{\partial \theta}\sin\phi \\ &= \sum_{\mu=1}^{n} P_\mu'(1)(a_{\mu 1}\cos\phi + b_{\mu 1}\sin\phi), \\ \lim_{\theta\to 0}((r\sin\theta)^{-1}U_\phi) &= -\frac{\partial A_1}{\partial \theta}\sin\phi + \frac{\partial B_1}{\partial \theta}\cos\phi \\ &= \sum_{\mu=1}^{n} P_\mu'(1)(-a_{\mu 1}\sin\phi + b_{\mu 1}\cos\phi), \end{aligned}$$

so that

$$\begin{aligned} K = \cos\alpha \sum_{\mu=1}^{n} \mu a_{\mu 0} &+ \sin\alpha\cos(\phi + \beta) \sum_{\mu=1}^{n} P_\mu'(1) a_{\mu 1} \\ &+ \sin\alpha\sin(\phi + \beta) \sum_{\mu=1}^{n} P_\mu'(1) b_{\mu 1}. \end{aligned} \tag{10}$$

3. According to the orthogonality property of the surface harmonics, we have

$$\int_0^{\pi}\int_{-\pi}^{+\pi} U(1,\theta',\phi')P_\mu(\cos\theta')\sin\theta'd\theta'd\phi' = \frac{4\pi}{2\mu+1}a_{\mu 0},$$

$$\int_0^{\pi}\int_{-\pi}^{+\pi} U(1,\theta',\phi')\sin\theta'P_\mu'(\cos\theta')\begin{matrix}\cos\phi'\\ \sin\phi'\end{matrix}\sin\theta'd\theta'd\phi' = \frac{4\pi}{2\mu+1}\cdot\begin{cases}P_\mu'(1)a_{\mu 1},\\ P_\mu'(1)b_{\mu 1},\end{cases}$$

so that

$$K = \frac{1}{4\pi}\int_0^{\pi}\int_{-\pi}^{+\pi} U(1,\theta',\phi')\left\{\cos\alpha\sum_{\mu=1}^{n}\mu(2\mu+1)P_\mu(\cos\theta')\right.$$
$$\left. +\sin\alpha\cos(\phi-\phi'+\beta)\sum_{\mu=1}^{n}(2\mu+1)\sin\theta'P_\mu'(\cos\theta')\right\}\sin\theta'd\theta'd\phi'. \tag{11}$$

In these integrals, $0\leqq\theta'\leqq\pi$, $-\pi\leqq\phi'\leqq+\pi$.

4. Now, we introduce the Fourier expansions

$$\frac{1}{4}\pi\sin\theta\sum_{\mu=1}^{n}\mu(2\mu+1)P_\mu(\cos\theta)$$
$$= \rho_1\cos\theta+\rho_2\cos 2\theta+\cdots+\rho_n\cos n\theta+\cdots,$$
$$\frac{1}{8}\pi\sin^2\theta\sum_{\mu=1}^{n}(2\mu+1)P_\mu'(\cos\theta) \tag{12}$$
$$= \sigma_1\sin\theta+\sigma_2\sin 2\theta+\cdots+\sigma_n\sin n\theta+\cdots.$$

Since

$$\int_0^{\pi}\sin\theta P_\mu(\cos\theta)\cos\nu\theta d\theta = \int_0^{\pi}\sin^2\theta P_\mu'(\cos\theta)\sin\nu\theta d\theta = 0, \qquad \mu>\nu,$$

the coefficients ρ_ν, σ_ν, $1\leqq\nu\leqq n$, depend only on ν; that is to say, for a fixed value of ν, they are independent of n. Consequently, we find from (11):

$$K = \frac{1}{2\pi^2}\int_{-\pi}^{+\pi}\int_{-\pi}^{+\pi} U(1,\theta',\phi')\left\{\cos\alpha\sum_{\nu=1}^{n}\rho_\nu\cos\nu\theta'\right.$$
$$\left. +2\sin\alpha\cos(\phi-\phi'+\beta)\sum_{\nu=1}^{n}\sigma_\nu\sin\nu\theta'\right\}d\theta'd\phi'. \tag{13}$$

Here both θ', ϕ' run from $-\pi$ to $+\pi$; we use the fact that

$$\int_{-\pi}^{+\pi} U(1,\theta',\phi')d\phi'$$

is a cosine polynomial of the nth degree, and also that both integrals

$$\int_{-\pi}^{+\pi} U(1, \theta', \phi') \begin{matrix} \cos \phi' \\ \sin \phi' \end{matrix} d\phi'$$

represent sine polynomials of the nth degree.

SECOND PART OF THE PROOF

The basic idea of this part is similar to that used in the proof of II (see the Introduction).

1. We write (13) in the form

$$(13') \quad K = \frac{1}{2\pi^2} \int_{-\pi}^{+\pi} \int_{-\pi}^{+\pi} U(1, \theta', \phi') \{ \cos \alpha \cdot \rho(\theta') + 2 \sin \alpha \cos (\phi - \phi' + \beta) \cdot \sigma(\theta') \} d\theta' d\phi',$$

where

$$(14) \qquad \rho(\theta') = \sum_{\nu=1}^{n} \rho_\nu \cos \nu\theta', \qquad \sigma(\theta') = \sum_{\nu=1}^{n} \sigma_\nu \sin \nu\theta'.$$

Let us put

$$(15) \quad \rho_n \cos \alpha \cos n\theta' + 2\sigma_n \sin \alpha \cos (\phi - \phi' + \beta) \sin n\theta' = R \cos (n\theta' - \delta),$$

where $R > 0$, δ real, and R, δ are independent of θ'; obviously, R and δ depend on $\alpha, \beta, \phi, \phi', n$. We have

$$(16) \qquad R = \{ \rho_n^2 \cos^2 \alpha + 4\sigma_n^2 \sin^2 \alpha \cos^2 (\phi - \phi' + \beta) \}^{1/2}.$$

Furthermore, for $1 \leqq \nu \leqq n-1$,

$$\begin{aligned} \rho_{n-\nu} \cos \alpha \cos (n - \nu)\theta' + 2\sigma_{n-\nu} \sin \alpha \cos (\phi - \phi' + \beta) \sin (n - \nu)\theta' \\ = (\rho_{n-\nu}/\rho_n) R \cos \delta \cos \{ (n\theta' - \delta) - (\nu\theta' - \delta) \} \\ + (\sigma_{n-\nu}/\sigma_n) R \sin \delta \sin \{ (n\theta' - \delta) - (\nu\theta' - \delta) \}. \end{aligned}$$

Adding to this expression

$$(\rho_{n-\nu}/\rho_n) R \cos \delta \cos \{ (n\theta' - \delta) + (\nu\theta' - \delta) \} - (\sigma_{n-\nu}/\sigma_n) R \sin \delta \sin \{ (n\theta' - \delta) + (\nu\theta' - \delta) \},$$

we obtain

$$2R \cos (n\theta' - \delta) \{ (\rho_{n-\nu}/\rho_n) \cos \delta \cos (\nu\theta' - \delta) - (\sigma_{n-\nu}/\sigma_n) \sin \delta \sin (\nu\theta' - \delta) \}.$$

Consequently, since $U(1, \theta', \phi')$ is a trigonometric polynomial of the nth degree in θ',

$$K = \frac{1}{2\pi^2}\int_{-\pi}^{+\pi}\int_{-\pi}^{+\pi} U(1, \theta', \phi')\cdot R\cos(n\theta' - \delta)$$

$$\cdot\left\{1 + 2\sum_{\nu=1}^{n-1}\frac{\rho_{n-\nu}}{\rho_n}\cos\delta\cos(\nu\theta' - \delta) - \frac{\sigma_{n-\nu}}{\sigma_n}\sin\delta\sin(\nu\theta' - \delta)\right\}d\theta' d\phi'. \tag{17}$$

2. We now replace $\cos(n\theta' - \delta)$ by

$$\frac{1}{4r}\left\{\frac{1 - r^2}{1 - 2r\cos(n\theta' - \delta) + r^2} - \frac{1 - r^2}{1 + 2r\cos(n\theta' - \delta) + r^2}\right\}, \qquad 0 < r < 1,$$

and obtain in the usual manner*

$$K = \lim_{r\to 1-0}\frac{1}{8\pi^2 r}\int_{-\pi}^{+\pi}\int_{-\pi}^{+\pi} U(1, \theta', \phi')$$

$$\cdot R\left\{\frac{1 - r^2}{1 - 2r\cos(n\theta' - \delta) + r^2} - \frac{1 - r^2}{1 + 2r\cos(n\theta' - \delta) + r^2}\right\}$$

$$\cdot\left\{1 + 2\sum_{\nu=1}^{n-1}\left(\frac{\rho_{n-\nu}}{\rho_n}\cos\delta\cos(\nu\theta' - \delta)\right.\right.$$

$$\left.\left. - \frac{\sigma_{n-\nu}}{\sigma_n}\sin\delta\sin(\nu\theta' - \delta)\right)\right\}d\theta' d\phi'$$

$$= \frac{1}{4\pi n}\int_{-\pi}^{+\pi}\sum{}' U(1, \theta', \phi')R\left\{1 + 2\sum_{\nu=1}^{n-1}\left(\frac{\rho_{n-\nu}}{\rho_n}\cos\delta\cos(\nu\theta' - \delta)\right.\right.$$

$$\left.\left. - \frac{\sigma_{n-\nu}}{\sigma_n}\sin\delta\sin(\nu\theta' - \delta)\right)\right\}d\phi'$$

$$- \frac{1}{4\pi n}\int_{-\pi}^{+\pi}\sum{}'' U(1, \theta', \phi')R\left\{1 + 2\sum_{\nu=1}^{n-1}\left(\frac{\rho_{n-\nu}}{\rho_n}\cos\delta\cos(\nu\theta' - \delta)\right.\right.$$

$$\left.\left. - \frac{\sigma_{n-\nu}}{\sigma_n}\sin\delta\sin(\nu\theta' - \delta)\right)\right\}d\phi'. \tag{18}$$

The summations $\sum', \sum''$ are extended over the modulo 2π incongruent values of θ' for which $n\theta' - \delta \equiv 0$ and π, respectively. (These particular values of θ' depend, of course, on $\alpha, \beta, \phi, \phi', n$.) For these values

$$1 + 2\sum_{\nu=1}^{n-1}\left(\frac{\rho_{n-\nu}}{\rho_n}\cos\delta\cos(\nu\theta' - \delta) - \frac{\sigma_{n-\nu}}{\sigma_n}\sin\delta\sin(\nu\theta' - \delta)\right)$$

$$= 1 + 2\sum_{\nu=1}^{n-1}\left(\frac{\rho_{n-\nu}}{\rho_n}\cos n\theta'\cos(n - \nu)\theta' + \frac{\sigma_{n-\nu}}{\sigma_n}\sin n\theta'\sin(n - \nu)\theta'\right) = M_n(\theta').$$

* Szegö, loc. cit., p. 64.

Since $\sigma_n = n$ (see (29)), this is the same trigonometric polynomial of order $2n-1$ as defined in (4). It depends only on n.

Consequently,

$$K = \frac{1}{4\pi n}\int_{-\pi}^{+\pi} \sum' U(1, \theta', \phi') R M_n(\theta') d\phi' - \frac{1}{4\pi n}\int_{-\pi}^{+\pi} \sum'' U(1, \theta', \phi') R M_n(\theta') d\phi'. \tag{19}$$

Here and in the next section, $\sum'$ and $\sum''$ have the meaning given above.

3. In the last part of the proof we are going to show that $M_n(\theta') > 0$ for all real values of θ'. Anticipating this result, we obtain from (19)

$$|K| \leqq \frac{1}{4\pi n}\int_{-\pi}^{+\pi} \sum' R M_n(\theta') d\phi' + \frac{1}{4\pi n}\int_{-\pi}^{+\pi} \sum'' R M_n(\theta') d\phi'. \tag{20}$$

But we easily see that the trigonometric polynomial $M_n(\theta')$ does not contain any term with $\cos n\theta'$ and $\sin n\theta'$, and also that its absolute term is 1. Hence

$$\frac{1}{n}\sum' M_n(\theta') = \frac{1}{n}\sum'' M_n(\theta') = 1, \tag{21}$$

so that on account of (16),

$$|K| \leqq \frac{1}{4\pi}\int_{-\pi}^{+\pi} R d\phi' + \frac{1}{4\pi}\int_{-\pi}^{+\pi} R d\phi' = \frac{1}{2\pi}\int_{-\pi}^{+\pi} (\rho_n^2 \cos^2\alpha + 4\sigma_n^2 \sin^2\alpha \cos^2\phi')^{1/2} d\phi'. \tag{22}$$

By means of Schwarz's inequality

$$K^2 \leqq \frac{1}{2\pi}\int_{-\pi}^{+\pi} (\rho_n^2 \cos^2\alpha + 4\sigma_n^2 \sin^2\alpha \cos^2\phi') d\phi' = \rho_n^2 \cos^2\alpha + 2\sigma_n^2 \sin^2\alpha \leqq \rho_n^2,$$

provided that $\rho_n^2 \geqq 2\sigma_n^2$. This is the case for $n \geqq 4$ (see (30)). For these values of n, the inequality $|K| \leqq \rho_n$ is proved.

4. The preceding argument, combined with (30), shows that $|K| < \rho_n$ holds provided that $\sin^2\alpha > 0$. Now let $\sin^2\alpha = 0$, $\cos\alpha = 1$; then $\delta = 0$, $R = \rho_n$. In this case (19) reads as follows:

$$K = \frac{\rho_n}{2n}\sum' U(\theta') M_n(\theta') - \frac{\rho_n}{2n}\sum'' U(\theta') M_n(\theta').$$

Here $U(\theta') = (2\pi)^{-1}\int_{-\pi}^{+\pi} U(1, \theta', \phi') d\phi'$, and the sums $\sum'$, $\sum''$ are extended

over the modulo 2π incongruent solutions of the equations $n\theta' \equiv 0$ and π, respectively. Taking (21) into consideration, we see that $K = \rho_n$ holds if and only if $U(\theta') = \cos n\theta'$ for the special values of θ' mentioned. Since the cosine polynomial $U(\theta') - \cos n\theta'$ vanishes for the values $\theta' = 0, \pi/n, 2\pi/n, \cdots, \pi$ (for which $\cos \theta'$ assumes $n+1$ distinct values), it must vanish identically, that is, $U(\theta) \equiv \cos n\theta$.

Similarly $K = -\rho_n$ holds if and only if $U(\theta) = -\cos n\theta$.

Considering the case $U(\theta) = \cos n\theta$, we write again $\theta = \theta' = k\pi/n$, k an integer. This furnishes

$$\frac{1}{2\pi}\int_{-\pi}^{+\pi} U(1, \theta', \phi')d\phi' = (-1)^k.$$

Since $|U(1, \theta', \phi')| \leqq 1$, we find $U(1, \theta', \phi') = (-1)^k$ identically in ϕ'. Consequently, by use of the notation (8), (9), we have for the values of θ' mentioned before, $A_\nu(1, \theta') = B_\nu(1, \theta') = 0$, $\nu = 1, 2, \cdots, n$, or if $\sin \theta' \neq 0$,

$$\sum_{\mu=\nu}^{n} a_{\mu\nu} P_\mu^{(\nu)}(\cos \theta') = \sum_{\mu=\nu}^{n} b_{\mu\nu} P_\mu^{(\nu)}(\cos \theta') = 0.$$

These equations hold in particular for $\theta' = \pi/n, 2\pi/n, \cdots, (n-1)\pi/n$, so that $A_\nu(1, \theta)$ and $B_\nu(1, \theta)$ must vanish identically for ν greater than or equal to 2, and $A_1(1, \theta) = \text{const.} \sin n\theta$, $B_\nu(1, \theta) = \text{const.} \sin n\theta$. Finally, $U(1, \theta, \phi) = \cos n\theta + c \sin n\theta \cos(\phi - \phi_0)$, c and ϕ_0 constant, and as a consequence of $|U| \leqq 1$ we find $c = 0$.

Calculation of ρ_n and σ_n

1. The constants ρ_n and σ_n involved in the previous considerations, can be calculated in various ways. A comparatively simple procedure is based on the well known formulas

$$\begin{aligned} \lim_{\epsilon \to +0} \{Q_n(u + i\epsilon) - Q_n(u - i\epsilon)\} &= -i\pi P_n(u), \\ \lim_{\epsilon \to +0} \{Q_n'(u + i\epsilon) - Q_n'(u - i\epsilon)\} &= -i\pi P_n'(u), \end{aligned} \tag{23}$$

where $-1 < u < +1$ and $Q_n(\xi)$ denotes Legendre's function of the second kind. Let $z = re^{i\theta}$, $0 < r < 1$, $0 < \theta < \pi$; then (23) can be written as follows:

$$\begin{aligned} P_n(\cos \theta) &= \frac{2}{\pi} \lim_{r \to 1-0} \Im Q_n(\tfrac{1}{2}(z + z^{-1})), \\ P_n'(\cos \theta) &= \frac{2}{\pi} \lim_{r \to 1-0} \Im Q_n'(\tfrac{1}{2}(z + z^{-1})), \end{aligned} \tag{24}$$

so that

$$
\begin{aligned}
&\frac{1}{4}\pi \sin\theta \sum_{\mu=1}^{n} \mu(2\mu+1)P_\mu(\cos\theta) \\
&\quad = \lim_{r\to 1-0} \frac{\pi}{4}\,\frac{z-z^{-1}}{2i}\,\frac{2}{\pi}\,\Im \sum_{\mu=1}^{n} \mu(2\mu+1)Q_\mu(\tfrac{1}{2}(z+z^{-1})) \\
&\quad = \lim_{r\to 1-0} \Re\left\{\frac{1-z^2}{4z}\sum_{\mu=1}^{n}\mu(2\mu+1)Q_\mu(\tfrac{1}{2}(z+z^{-1}))\right\}, \\
&\frac{1}{8}\pi \sin^2\theta \sum_{\mu=0}^{n} (2\mu+1)P_\mu'(\cos\theta) \\
&\quad = \lim_{r\to 1-0} \frac{\pi}{8}\left(\frac{z-z^{-1}}{2i}\right)^2\frac{2}{\pi}\,\Im \sum_{\mu=0}^{n}(2\mu+1)Q_\mu'(\tfrac{1}{2}(z+z^{-1})) \\
&\quad = -\lim_{r\to 1-0} \Im\left\{\frac{(1-z^2)^2}{16z^2}\sum_{\mu=0}^{n}(2\mu+1)Q_\mu'(\tfrac{1}{2}(z+z^{-1}))\right\}.
\end{aligned}
\tag{25}
$$

Now $Q_\mu((z+z^{-1})/2)$ is a power series of z starting with $z^{\mu+1}$; therefore $z^{-2}Q_\mu'((z+z^{-1})/2)$ is a power series starting with z^μ. Adding to the sums on the right-hand side of (25) the corresponding terms with $\mu>n$, we do not influence the coefficients of $z, z^2, \cdots, z^n$. From this, we conclude that the n first coefficients of the expansions (12) agree with the corresponding coefficients of

$$
\begin{aligned}
f(z) &= \frac{1-z^2}{4z}\sum_{\mu=1}^{\infty}\mu(2\mu+1)Q_\mu(\tfrac{1}{2}(z+z^{-1})), \\
g(z) &= -\frac{(1-z^2)^2}{16z^2}\sum_{\mu=0}^{\infty}(2\mu+1)Q_\mu'(\tfrac{1}{2}(z+z^{-1})).
\end{aligned}
\tag{26}
$$

In particular, ρ_n and σ_n are the coefficients of z^n in the power series expansions of the functions (26), regular for $|z|<1$.

2. In order to compute the functions (26), we start from the classical formula

$$
\begin{aligned}
\frac{1}{\xi-1} &= \sum_{\mu=0}^{n}(2\mu+1)Q_\mu(\xi) + \sum_{\mu=n+1}^{\infty}(2\mu+1)Q_\mu(\xi) \\
&= \sum_{\mu=0}^{n}(2\mu+1)Q_\mu(\xi) + \frac{n+1}{\xi-1}\{Q_n(\xi)-Q_{n+1}(\xi)\} \\
&= \sum_{\mu=0}^{n}(2\mu+1)Q_\mu(\xi) + r_{n+1}(\xi).
\end{aligned}
\tag{27}
$$

Here $\xi=(z+z^{-1})/2$, $|z|<1$, is in the complex plane cut along the segment $[-1, +1]$. Now

$$\sum_{\mu=1}^{\infty} \mu(2\mu+1)Q_\mu(\xi) = \sum_{\mu=1}^{\infty} \mu(r_\mu(\xi) - r_{\mu+1}(\xi)) = r_1(\xi) + r_2(\xi) + r_3(\xi) + \cdots$$

$$= \frac{1}{\xi - 1} \sum_{\mu=1}^{\infty} \mu(Q_{\mu-1}(\xi) - Q_\mu(\xi)) = \frac{1}{\xi - 1} \sum_{\mu=0}^{\infty} Q_\mu(\xi),$$

so that from the well known integral representation of $Q_\mu(\xi)$, by use of the generating function of $P_\mu(t)$,

$$\sum_{\mu=1}^{\infty} \mu(2\mu+1)Q_\mu(\xi) = \frac{1}{\xi - 1} \sum_{\mu=0}^{\infty} \frac{1}{2} \int_{-1}^{+1} \frac{P_\mu(t)}{\xi - t}\, dt = \frac{2^{-3/2}}{\xi - 1} \int_{-1}^{+1} \frac{(1-t)^{-1/2}}{\xi - t}\, dt$$

$$= \frac{2^{-1/2}}{(\xi - 1)^{3/2}} \arctan\left(\frac{2}{\xi - 1}\right)^{1/2}$$

$$= \frac{2z^{3/2}}{(1-z)^3} \arctan\left(\frac{2z^{1/2}}{1-z}\right)$$

$$= \frac{4z^{3/2}}{(1-z)^3} \arctan(z^{1/2}).$$

Consequently,

$$f(z) = \frac{1+z}{(1-z)^2} z^{1/2} \arctan(z^{1/2}) = \frac{1+z}{(1-z)^2} \frac{z^{1/2}}{2i} \log \frac{1+iz^{1/2}}{1-iz^{1/2}}$$

$$= \sum_{\nu=0}^{\infty} (2\nu+1)z^\nu \cdot \sum_{\nu=1}^{\infty} \frac{(-1)^{\nu-1}}{2\nu-1} z^\nu,$$

(28)

$$\rho_n = \sum_{\nu=1}^{n} \frac{(-1)^{\nu-1}}{2\nu-1} (2n - 2\nu + 1)$$

$$= 2n\left(\frac{1}{1} - \frac{1}{3} + \cdots + \frac{(-1)^{n-1}}{2n-1}\right) + \begin{cases} 0, & n \text{ even}, \\ -1, & n \text{ odd}, \end{cases}$$

which agrees with the expression in the right-hand member of the inequality (3).

3. A simpler argument furnishes from (27)

$$\sum_{\mu=0}^{\infty} (2\mu+1)Q'_\mu(\xi) = -(\xi - 1)^{-2} = -4z^2(1-z)^{-4},$$

(29)

$$g(z) = \frac{1}{4}\left(\frac{1+z}{1-z}\right)^2, \quad \sigma_n = n.$$

4. It can be easily seen that

(30) $$\rho_n > 2^{1/2}\sigma_n \qquad \text{for } n \geqq 4.$$

Indeed, we write

$$\frac{\rho_n}{2n} = 1 - \frac{1}{3} + \cdots + \frac{(-1)^{n-1}}{2n-1} + \frac{\epsilon_n}{2n},$$

where $\epsilon_n = 0$ or -1, according as n is even or odd. If n is even, we find, $n \geqq 4$,

$$\frac{\rho_n}{2n} \geqq 1 - \frac{1}{3} + \frac{1}{5} - \frac{1}{7} = \frac{76}{105} > 2^{-1/2}.$$

If n is odd, $n \geqq 5$,

$$\frac{\rho_n}{2n} = \frac{\rho_{n-1}}{2(n-1)} + \frac{1}{2n-1} - \frac{1}{2n} > \frac{\rho_{n-1}}{2(n-1)} > 2^{-1/2}.$$

Apparently, (30) does not hold for $n=1, 2, 3$. Indeed $\rho_1=1$, $\rho_2=8/3$, $\rho_3=21/5$, $\sigma_1=1$, $\sigma_2=2$, $\sigma_3=3$.

Positivity of $M_n(\theta)$

Let $n \geqq 2$. The remaining part of the proof is devoted to the discussion of the trigonometric polynomial (4), that is, of

$$M_n(\theta) = 2\sum_{\nu=0}^{n-1}{}^{*}\left(\frac{\rho_{n-\nu}}{\rho_n}\cos n\theta \cos (n-\nu)\theta + \frac{n-\nu}{n}\sin n\theta \sin (n-\nu)\theta\right),$$

where the symbol $\sum^*$ indicates that the term $\nu=0$ must be multiplied by $1/2$. We want to show that $M_n(\theta) > 0$.

1. Let $0 < \theta < \pi$. For the function $f(z)$ in (28), we have $f(0)=0$, so that

$$\begin{aligned} 2\sum_{\nu=0}^{n-1}{}^{*}\rho_{n-\nu}\cos (n-\nu)\theta &= \frac{1}{2\pi i}\int \frac{f(z)}{z^{n+1}}\left\{2\sum_{\nu=0}^{n-1}{}^{*} z^{\nu}\cos (n-\nu)\theta\right\} dz \\ &= \frac{1}{2\pi i}\int \frac{f(z)}{z^{n+1}}\left\{\cos n\theta + 2\sum_{\nu=1}^{\infty} z^{\nu}\cos (n-\nu)\theta\right\} dz \qquad (31) \\ &= \frac{1}{2\pi i}\int \frac{f(z)}{z^{n+1}}\,\frac{(1-z^2)\cos n\theta + 2z\sin n\theta \sin\theta}{1 - 2z\cos\theta + z^2}\, dz. \end{aligned}$$

In an arbitrary domain not containing $z=-1$, the function $f(z)$ has the only singularity $z=+1$; an additional singularity of the last integrand is at $z=e^{\pm i\theta}$. The integrand being single-valued around $z=+1$, $e^{\pm i\theta}$, we can choose as a contour the following: A "large" arc $|z|=R$, $|\arg z| \leqq \pi-\epsilon$, and (approximately) the segment $-R$, -1 of the negative real axis, counting twice and encircling $z=-1$ in the negative sense. The residues at $z=+1$, $e^{\pm i\theta}$ have to be taken into consideration.

Now, since $f(z)=O(|z|^{-1/2})$ as $z\to\infty$ (compare the second form of $f(z)$ given in (28)), the contribution of the large arc can be disregarded. The residues furnish

$$(32)\qquad \begin{aligned} &-\frac{d}{dz}\left\{(1+z)z^{-n-1/2}\arctan(z^{1/2})\,\frac{(1-z^2)\cos n\theta+2z\sin n\theta\sin\theta}{1-2z\cos\theta+z^2}\right\}_{z=+1}\\ &-2\Re\left\{e^{-i(n+1)\theta}f(e^{i\theta})\,\frac{(1-e^{2i\theta})\cos n\theta+2e^{i\theta}\sin n\theta\sin\theta}{e^{i\theta}-e^{-i\theta}}\right\}. \end{aligned}$$

The first term is

$$\begin{aligned} &(-\pi/4+(2n+1)\pi/4-1/2)\,\frac{2\sin n\theta\sin\theta}{2(1-\cos\theta)}-\frac{\pi}{2}\,\frac{-2\cos n\theta}{2(1-\cos\theta)}\\ &\qquad=\frac{(n\pi-1)\sin n\theta\sin\theta+\pi\cos n\theta}{4\sin^2(\theta/2)}. \end{aligned}$$

For the calculation of the second term, we note that for $0<\theta<\pi$

$$(33)\qquad \begin{aligned} f(e^{i\theta})&=\frac{1+e^{i\theta}}{(1-e^{i\theta})^2}\,\frac{e^{i\theta/2}}{2i}\left\{\log\cot\left[(\theta+\pi)/4\right]+i\pi/2\right\}\\ &=\frac{\cos(\theta/2)}{4\sin^2(\theta/2)}\left\{i\log\cot\left[(\theta+\pi)/4\right]-\pi/2\right\}, \end{aligned}$$

so that the second term of (32) is

$$-2\Re\left\{e^{-i(n+1)\theta}f(e^{i\theta})(-e^{i(n+1)\theta})\right\}=2\Re f(e^{i\theta})=\frac{-\pi\cos(\theta/2)}{4\sin^2(\theta/2)}.$$

Finally, we consider

$$(34)\qquad \frac{1}{2\pi i}\int_{-\infty}^{(-1,-)}\frac{f(z)}{z^{n+1}}\,\frac{(1-z^2)\cos n\theta+2z\sin n\theta\sin\theta}{1-2z\cos\theta+z^2}\,dz.$$

On the "upper" border of the line $-\infty$, -1 we have, $z=-t$, $1<t<+\infty$,

$$f(z)=\frac{1-t}{(1+t)^2}\,\frac{t^{1/2}}{2}\left\{\log\left(\frac{t^{1/2}-1}{t^{1/2}+1}\right)+i\pi\right\},$$

whereas on the "lower" border, $i\pi$ must be replaced by $-i\pi$. Consequently, (34) will be

$$\frac{(-1)^n}{2}\int_1^\infty\frac{t-1}{(t+1)^2}\,t^{-n-1/2}\,\frac{(1-t^2)\cos n\theta-2t\sin n\theta\sin\theta}{1+2t\cos\theta+t^2}\,dt,$$

so that

$$2\sum_{\nu=0}^{n-1}{}^{*}\rho_{n-\nu}\cos(n-\nu)\theta = \frac{(n\pi-1)\sin n\theta\sin\theta+\pi\cos n\theta-\pi\cos(\theta/2)}{4\sin^2(\theta/2)}$$

$$+\frac{(-1)^n}{2}\int_1^\infty \frac{t-1}{(t+1)^2}\,t^{-n-1/2}\,\frac{(1-t^2)\cos n\theta-2t\sin n\theta\sin\theta}{1+2t\cos\theta+t^2}\,dt. \tag{35}$$

2. According to a well known formula,

$$1+2\sum_{\nu=1}^{n-1}\cos\nu\theta+\cos n\theta = \frac{\sin(2n-1)(\theta/2)}{2\sin(\theta/2)}+\frac{\sin(2n+1)(\theta/2)}{2\sin(\theta/2)}$$

$$=\sin n\theta\cot(\theta/2),$$

so that

$$2\sum_{\nu=0}^{n-1}{}^{*}(n-\nu)\sin(n-\nu)\theta = 2\sum_{\nu=1}^{n-1}\nu\sin\nu\theta+n\sin n\theta$$

$$=-n\cos n\theta\cot(\theta/2)+\frac{\sin n\theta}{2\sin^2(\theta/2)}.$$

This leads, in view of (35), to the following important representation:

$$\rho_n M_n(\theta) = \frac{(n\pi-1)\sin n\theta\sin\theta+\pi\cos n\theta-\pi\cos(\theta/2)}{4\sin^2(\theta/2)}\cos n\theta$$

$$+\frac{(-1)^{n-1}}{2}\cos n\theta\int_1^\infty\frac{t-1}{(t+1)^2}\,t^{-n-1/2}\Re\left\{e^{in\theta}\frac{1-t^{-1}e^{i\theta}}{1+t^{-1}e^{i\theta}}\right\}dt \tag{36}$$

$$-\tfrac{1}{2}\rho_n\sin 2n\theta\cot(\theta/2)+\frac{\rho_n}{2n}\,\frac{\sin^2 n\theta}{\sin^2(\theta/2)}.$$

3. Now, we first establish the inequality

$$\left|\frac{(-1)^{n-1}}{2}\cos n\theta\int_1^\infty\frac{t-1}{(t+1)^2}\,t^{-n-1/2}\Re\left\{e^{in\theta}\frac{1-t^{-1}e^{i\theta}}{1+t^{-1}e^{i\theta}}\right\}dt\right|<\frac{1}{4n}. \tag{37}$$

Indeed,

$$\left|\Re\left\{e^{in\theta}\frac{1-t^{-1}e^{i\theta}}{1+t^{-1}e^{i\theta}}\right\}\right|\leqq\frac{1+t^{-1}}{1-t^{-1}},$$

and

$$\frac{1}{2}\int_1^\infty\frac{t^{-n-1/2}}{t+1}\,dt<\frac{1}{4}\int_1^\infty t^{-n-1}dt=\frac{1}{4n}.$$

4. Next, we proceed to a discussion of the quantities ρ_n. If s_n denotes the

nth partial sum of the series $1/1-1/3+1/5-\cdots$, and $\epsilon_n=0$ or -1 according as n is even or odd, we have

$$\begin{aligned}\rho_n &= 2ns_n+\epsilon_n = 2n\left(\pi/4-\sum_{\nu=n+1}^{\infty}\frac{(-1)^{\nu-1}}{2\nu-1}\right)+\epsilon_n \\ &= n\pi/2+\epsilon_n+(-1)^{n+1}2n\left\{\left(\frac{1}{2n+1}-\frac{1}{2n+3}\right)\right. \\ &\quad \left.+\left(\frac{1}{2n+5}-\frac{1}{2n+7}\right)+\cdots\right\}.\end{aligned}$$

We compare the last series with

$$\frac{1}{2}\left(\frac{1}{2n}-\frac{1}{2n+4}\right)+\frac{1}{2}\left(\frac{1}{2n+4}-\frac{1}{2n+8}\right)+\cdots=\frac{1}{4n}.$$

Since

$$\begin{aligned}0 &< \frac{1}{2}\left(\frac{1}{2n}-\frac{1}{2n+4}\right)-\left(\frac{1}{2n+1}-\frac{1}{2n+3}\right) \\ &= \frac{6}{2n(2n+1)(2n+3)(2n+4)}<\frac{3}{8n^4},\end{aligned}$$

we find

$$\rho_n = n\pi/2+\epsilon_n+(-1)^{n+1}/2+\delta_n = \tfrac{1}{2}(n\pi-1)+\delta_n, \tag{38}$$

where

$$\begin{aligned}|\delta_n| &< \frac{3n}{4}\left(\frac{1}{n^4}+\frac{1}{(n+2)^4}+\frac{1}{(n+4)^4}+\cdots\right) \\ &< \frac{3}{4}\frac{1}{n^3}+\frac{3}{8}n\int_n^{\infty}\frac{dx}{x^4}=\frac{3}{4}\frac{1}{n^3}+\frac{1}{8}\frac{1}{n^2}.\end{aligned} \tag{39}$$

Incidentally, $\operatorname{sgn}\delta_n=(-1)^n$.

5. Returning again to (36), we obtain

$$\begin{aligned}\rho_n M_n(\theta) &= \frac{(n\pi-1)\sin n\theta\sin\theta+\pi\cos n\theta-\pi\cos(\theta/2)}{4\sin^2(\theta/2)}\cos n\theta \\ &\quad -\tfrac{1}{4}(n\pi-1)\sin 2n\theta\cot(\theta/2)+\frac{n\pi-1}{4n}\frac{\sin^2 n\theta}{\sin^2(\theta/2)}+\delta_n' \\ &= \frac{\pi(1-\cos(\theta/2)\cos n\theta)-n^{-1}\sin^2 n\theta}{4\sin^2(\theta/2)}+\delta_n',\end{aligned} \tag{40}$$

where

$$(41)\qquad \begin{aligned} |\delta_n'| &< \frac{1}{4n} + |\delta_n|\left(\frac{1}{2}|\sin 2n\theta|\cot(\theta/2) + \frac{1}{2n}\frac{\sin^2 n\theta}{\sin^2(\theta/2)}\right) \\ &< \frac{1}{4n} + |\delta_n|\left(\frac{1}{2}\cdot 2n\theta\cdot\frac{2}{\theta} + \frac{1}{2n}\cdot 4n^2\right) = \frac{1}{4n} + 4n|\delta_n|, \end{aligned}$$

or, according to (39),

$$(42)\qquad |\delta_n'| < 3/(4n) + 3/n^2.$$

6. In order to prove the positivity of (40), we distinguish two cases.

(a) Let $(1/2)n\pi\cos(\theta/2) < 1$. Then

$$(43)\qquad \pi(1 - \lambda\cos(\theta/2)) - n^{-1}(1 - \lambda^2),$$

where λ varies between -1 and $+1$, has a minimum at $\lambda = (1/2)n\pi\cos(\theta/2)$ which is $\pi - n^{-1} - (1/4)n\pi^2\cos^2(\theta/2)$. Therefore, we only have to show that

$$\frac{\pi - n^{-1} - (1/4)n\pi^2\cos^2(\theta/2)}{4\sin^2(\theta/2)} > \frac{3}{4n} + \frac{3}{n^2}.$$

Since $\pi - n^{-1} - n\pi^2/4 < 0$, the expression on the left-hand side is a decreasing function of $\cos(\theta/2)$, so that its minimum is attained for $\cos(\theta/2) = 2/n\pi$. But

$$\frac{\pi - n^{-1} - n^{-1}}{4(1 - 4/n^2\pi^2)} = \frac{1}{4}\pi\left(1 + \frac{2}{n\pi}\right)^{-1} > \frac{3}{4n} + \frac{3}{n^2};$$

this inequality is true for $n \geqq 3$.

(b) Let $(1/2)n\pi\cos(\theta/2) > 1$. Then (43) is decreasing and its minimum is reached for $\lambda = +1$, which is $\pi(1 - \cos(\theta/2))$. Therefore, it is sufficient to show that

$$\frac{\pi(1 - \cos(\theta/2))}{4\sin^2(\theta/2)} = \frac{\pi}{4(1 + \cos(\theta/2))} \geqq \frac{\pi}{8} > \frac{3}{4n} + \frac{3}{n^2},$$

which is true for $n \geqq 4$.

It remains to show that for $n = 2$ and $n = 3$, we have $\pi/8 > |\delta_n'|$, or using (41), $\pi/8 > 1/(4n) + 4n|\delta_n|$. Since $\rho_2 = 8/3$, $\rho_3 = 21/5$, we find from (38) $0 < \delta_2 = 19/6 - \pi$, $0 < -\delta_3 = 3\pi/2 - 47/10$, so that the statements are $\pi > 1 + 64(19/6 - \pi)$, $\pi > 2/3 + 96(3\pi/2 - 47/10)$. Both inequalities can easily be verified.

STANFORD UNIVERSITY,
STANFORD UNIVERSITY, CALIF.

ON MEASURABLE STOCHASTIC PROCESSES*

BY

WARREN AMBROSE

In recent years probability theory has been formulated mathematically as measure theory; in the case of stochastic processes depending upon a continuous parameter the measures considered are defined on certain subspaces of the space of all functions of a real variable.† This formulation of stochastic processes depending upon a continuous parameter gives rise to certain measurability problems, and it is with these measurability problems that this paper is concerned. In particular we shall be concerned with conditions under which there will exist what Doob has called a measurable stochastic process.‡ In §1 we give the necessary mathematical formulation of the notion of a stochastic process. In §2 we obtain general conditions for the existence of a measurable process, while in §3 we use the results of §2 to obtain conditions upon the conditional probability functions which are necessary and sufficient for the existence of a measurable process. In §4 we prove a theorem which is essentially due to W. Doeblin concerning the existence of a special sort of measurable process in case the conditional probabilities satisfy certain regularity conditions.

1. **Mathematical formulations.** We shall denote by Ω the space of all real-valued functions of a real variable. We introduce a topology on Ω by defining neighborhoods as follows: if $t_1, \cdots, t_n$ is any finite set of real numbers and if $a_1, \cdots, a_n$ and $b_1, \cdots, b_n$ are sets of real numbers satisfying $-\infty \leqq a_i < b_i \leqq \infty$ $(i=1, \cdots, n)$, then the set of elements $x(t)$ of Ω which satisfy $a_i < x(t_i) < b_i$ $(i=1, \cdots, n)$ is a neighborhood. Next we consider a probability measure $P(M)$, defined on the Borel field of sets determined by the collection of neighborhoods;§ we shall suppose the domain of definition of the measure $P(M)$ to be so extended that if $P(M)=0$ for a certain set M, then $P(N)$ is defined for every subset N of M. The sets for which $P(M)$ is defined will be called P-measurable. If N is any set in Ω, we define its outer P-meas-

* Presented to the Society, September 8, 1939, under the title *Some properties of measurable stochastic processes*; received by the editors June 5, 1939.

† See [2] and [5]. Numbers in square brackets refer to the bibliography at the end of this paper.

‡ See [2, p. 113].

§ The Borel field of sets determined by a given collection of sets is the smallest collection which contains the sets of the given collection and which has the further properties: (1) if M belongs to the collection, then so does the complement of M, (2) if $M_1, \cdots, M_n, \cdots$ belong to the collection, then so do $\sum_{i=1}^{\infty} M_i$ and $\prod_{i=1}^{\infty} M_i$. A measure is a completely additive nonnegative set-function; a probability measure is a measure for which $P(\Omega)=1$.

ure, denoted by $\overline{P(N)}$, to be the lower bound of numbers of the form $P(M)$, where M ranges through the P-measurable sets including N.

DEFINITION. *Let Ω' be a subset of Ω for which $\overline{P(\Omega')}=1$. If M' is the intersection of Ω' with a P-measurable set M, then we define the probability measure $P'(M')$ by $P'(M')=P(M)$.* The space Ω' together with the probability measure $P'(M')$ is a stochastic process.*†

For brevity we shall sometimes speak as though the space Ω' were itself the stochastic process; a reading of the context in which the term occurs will eliminate any ambiguity of meaning. We shall sometimes use the symbol ω and sometimes the symbol $x(t)$ to denote a point of Ω, using the latter symbol when we wish to emphasize the fact that a point of Ω is a function of a real variable. If ω is the function $x(t)$ and if s is a real number, then the symbol $x_s(\omega)$ will denote the value of $x(t)$ for $t=s$; this function $x_s(\omega)$ is fundamental in the measurability conditions that follow.

Let T denote the space of real numbers t; let $T\times\Omega'$ denote the product space of T with a space Ω'; then if t-measure is taken as Lebesgue measure, we can define a measure in this product space so that the measure of the direct product of a Lebesgue measurable set and a P'-measurable set is the product of their measures.‡ It is readily seen that a set in $T\times\Omega'$ is measurable if and only if it is the intersection of $T\times\Omega'$ with a set which is measurable in $T\times\Omega$.

DEFINITION. *The stochastic process consisting of Ω' with a P'-measure on it is measurable if $x_s(\omega)$ is measurable on $T\times\Omega'$.*

The measurability of this function $x_s(\omega)$ is of fundamental importance in theorems about stochastic processes. For example, it makes possible certain applications of the Birkhoff ergodic theorem.

2. **A general condition for the existence of a measurable stochastic process.** In this section we shall obtain a necessary and sufficient condition (on a P-measure) that Ω contain a measurable stochastic process. We shall then use this condition to derive an important necessary and sufficient condition, due to Kolmogoroff, for the existence of a measurable process.

LEMMA 1. *Let x-space be any space on which a measure is defined, and let E be any set lying in x-space. Suppose that $f(x)$ is a real-valued function, defined*

* It is shown in [2] that the value of the function $P'(M')$ is independent of the particular choice of the set M used in representing M' as the intersection of Ω' and a P-measurable set.

† This definition (as well as the definition of a measurable stochastic process which follows) is that given in [2]; see pp. 109–110 and p. 113. Also we remark that this is not the most general definition of a stochastic process but only a definition of the type of stochastic process to be studied in this paper.

‡ For a discussion of measures in product spaces see [6].

on E, with the property that for every real number k, the x-set (lying in E) for which $f(x)>k$ is the intersection of E with a measurable x-set. Then $f(x)$ can be so defined for x in the complement of E that it will become a measurable x-function.

Proof. By hypothesis we know that for every rational number r there is a measurable x-set $N(r)$, such that

$$_x[f(x) > r] = N(r) \cdot E.^*$$

We now define the measurable x-sets $M(r)$ by $M(r) = \sum_{r' \geq r} N(r')$, and it is obvious that

$$_x[f(x) > r] = M(r) \cdot E.$$

We define $f(x)$ for x lying in the complement of E as follows:

$$f(x) = \text{u.b.}_{x \in M(r)} r \quad \text{for} \quad x \in \sum_r M(r)$$

$$= 0 \qquad\qquad \text{for} \quad x \notin \sum_r M(r).$$

It is readily verified that $f(x)$ becomes a measurable x-function when so extended.

THEOREM 1. *The space Ω contains a measurable stochastic process if and only if P-measure is so defined that there is a measurable function on $T \times \Omega$ which for each fixed t equals $x_t(\omega)$ at almost all ω-points.*

Proof. We shall first show that if there is a measurable stochastic process, then there is a function $f(t, \omega)$ of the sort mentioned in this theorem. Suppose that Ω' is a measurable stochastic process; since the function $x_t(\omega)$ is measurable on $T \times \Omega'$ and since this means that for every real number k the set in $T \times \Omega'$ for which $x_t(\omega) > k$ is the intersection of $T \times \Omega'$ with a measurable set in $T \times \Omega$, we can apply Lemma 1 to this function (taking the x-space mentioned in that lemma to be $T \times \Omega$ and taking $T \times \Omega'$ for the set E). This lemma tells us that there is a measurable function on $T \times \Omega$, $g(t, \omega)$, which equals $x_t(\omega)$ on $T \times \Omega'$. Now for fixed t the function $g(t, \omega)$ is a measurable ω-function except possibly for a t-set whose measure is zero. We define the function $f(t, \omega)$ to be equal to $g(t, \omega)$, except for t in this exceptional set; for these exceptional t-values we define $f(t, \omega)$ to be equal to $x_t(\omega)$. Then the function $f(t, \omega)$ is measurable on $T \times \Omega$ and is also P-measurable for each value of t. We see that for each t the functions $f(t, \omega)$ and $x_t(\omega)$ are equal at almost all ω-points since they are both P-measurable and they are equal on Ω'.

* The symbol $_x[f(x) > r]$ denotes the set of x-points for which $f(x) > r$.

Next we shall show that if there is a measurable function on $T \times \Omega$, as mentioned in the theorem, then Ω contains a measurable stochastic process. Doob has shown* that if $f(t, \omega)$ is a measurable (t, ω)-function and if $\varphi_n(t)$ is defined by $\varphi_n(t) = k2^{-n}$ for $(k-1)2^{-n} < t \leqq k2^{-n}$, then there is a real number c and an increasing sequence of integers $\{a_n\}$ such that

$$\lim_{n \to \infty} f(\varphi_{a_n}(t) + c, \omega) = f(t + c, \omega)$$

at almost all (t, ω)-points. This implies that for almost all t

$$P\left(\lim_{n \to \infty} f(\varphi_{a_n}(t) + c, \omega) = f(t + c, \omega)\right) = 1$$

and hence (since the functions $f(t, \omega)$ and $x_t(\omega)$ are equal, for fixed t, at almost all ω-points) that for almost all t

$$P\left(\lim_{n \to \infty} x(\varphi_{a_n}(t) + c) = x(c + t)\right) = 1,$$

and hence, according to a theorem of Doob,† Ω contains a measurable stochastic process.

The next theorem which we shall prove enables us to obtain from Theorem 1 a condition found by Kolmogoroff to be necessary and sufficient for the existence of a measurable stochastic process; first, however, we introduce a few preliminary notions.

If $f(\omega)$ and $g(\omega)$ are P-measurable functions, we define the distance between them to be

$$\int_\Omega |\arctan f(\omega) - \arctan g(\omega)| \, dP.$$

Now let $f(t, \omega)$ be a (t, ω)-function which is P-measurable for each t. We consider the related function $\Phi(t)$ whose value for each t is the ω-function $f(t, \omega)$. Using the above definition of distance in the space of P-measurable functions, we see that the range of $\Phi(t)$ is a metric space. We shall say that $\Phi(t)$ is measurable if the origin of any sphere in its range space is Lebesgue measurable.

THEOREM 2. *$\Phi(t)$ is approximately continuous almost everywhere if and only if there is a measurable function on $T \times \Omega$ which for each t equals $f(t, \omega)$ at almost all ω-points.*‡

Proof. Suppose first that $\Phi(t)$ is approximately continuous almost every-

* See [2, p. 113].
† See [2, p. 115].
‡ This theorem is a generalization of a theorem due to Doob; see [4, p. 757].

where; from a lemma of Doob* it follows that there is a t-set T_0 of measure zero such that the range of $\Phi(t)$ for $t \notin T_0$ is separable. This means that there is a sequence of real numbers, $\{\alpha_n\}$, such that the sequence of ω-functions $\{f(\alpha_n, \omega)\}$ is dense in the family of ω-functions $\{f(t, \omega)\}$, $t \notin T_0$. We can find for each function of this sequence a separable collection† of P-measurable sets with respect to which it is measurable. Hence it follows that there is a single separable collection of P-measurable sets with respect to which every function of the sequence $\{f(\alpha_n, \omega)\}$ is measurable. Then, since every member of the family $\{f(t, \omega)\}$, $t \notin T_0$, can be approximated by members of this sequence, it follows that every member of the family $\{f(t, \omega)\}$, $t \notin T_0$, is measurable with respect to this separable collection of sets.

Because we are assuming $\Phi(t)$ to be approximately continuous almost everywhere, it follows immediately that for any P-measurable set M, the t-function $\int_M \arctan f(t,\omega) dP$ is approximately continuous almost everywhere and hence is measurable. This fact and the fact established in the preceding paragraph together imply, by Theorem 1 of [4], that there is a measurable function on $T \times \Omega$ which equals $f(t, \omega)$, for each fixed t, at almost all ω-points.

We shall now show that if there is a measurable function on $T \times \Omega$ which is so related to $f(t, \omega)$, then $\Phi(t)$ is approximately continuous almost everywhere. We shall show this by showing that $\Phi(t)$ is measurable and that there is a t-set T_0 of measure zero such that the range of $\Phi(t)$ for $t \notin T_0$ is separable. To show that $\Phi(t)$ is measurable we must show that for any sphere S in the space of P-measurable functions the t-set for which $\Phi(t)$ belongs to S is measurable; that is, we must show that if $g(\omega)$ is any P-measurable function and k is any real number, then the t-set for which

$$\int |\arctan f(t, \omega) - \arctan g(\omega)| \, dP < k$$

is measurable. If $f_0(t, \omega)$ is the function which is measurable on $T \times \Omega$ and equals $f(t, \omega)$, for each fixed t, at almost all ω-points, then the function $|\arctan f_0(t, \omega) - \arctan g(\omega)|$ will be a measurable (t, ω)-function so that, by Fubini's theorem,

$$\int |\arctan f(t, \omega) - \arctan g(\omega)| \, dP \equiv \int |\arctan f_0(t, \omega) - \arctan g(\omega)| \, dP$$

* See [4 p. 755, Lemma 1]. The portion of the lemma which we use can be stated (where $\Phi(t)$ still represents the ω-function $f(t, \omega)$): $\Phi(t)$ is approximately continuous almost everywhere if and only if $\Phi(t)$ is measurable and there is a t-set T_0 of measure zero such that the range of $\Phi(t)$, for $t \notin T_0$, is a separable space.

† A collection of P-measurable sets $\{M_\alpha\}$ with characteristic function $\{\psi_\alpha(\omega)\}$ is said to be separable if the ω-functions $\{\psi_\alpha(\omega)\}$ form a separable space with the metric defined above.

is a measurable t-function; hence the t-set defined by the above inequality is measurable.

By Theorem 1 of [4] we see that there is a t-set T_0 of measure zero and a separable collection F of P-measurable sets such that for $t \notin T_0$, $f(t, \omega)$ is measurable with respect to F. This means that there exists a sequence of P-measurable sets $\{M_n\}$ with characteristic functions $\{\psi_n(\omega)\}$, such that if N is any set belonging to F, with characteristic function $\psi(\omega)$, then there is a subsequence $\{\psi_{\alpha_n}(\omega)\}$ of $\{\psi_n(\omega)\}$ such that

$$\int |\arctan \psi_{\alpha_n}(\omega) - \arctan \psi(\omega)| \, dP \to 0, \qquad n \to \infty.$$

It is readily seen that the collection of linear combinations, with rational coefficients, of these functions $\{\psi_n(\omega)\}$ is dense in the family $\{f(t, \omega)\}$, $t \notin T_0$, and hence that the range of $\Phi(t)$, $t \notin T_0$, is separable.

We are now able to obtain the Kolmogoroff theorem for the existence of a measurable process:

THEOREM 3. *Ω contains a measurable stochastic process if and only if P-measure is so defined that for every $\epsilon > 0$ and almost all t*

$$P(|x(t+h) - x(t)| > \epsilon) \to 0$$

*as h approaches zero on an h-set, which may depend upon t but not upon ϵ, having density 1 at $h = 0$.**

Before proving this theorem we restate it as follows: Ω contains a measurable stochastic process if and only if P-measure is so defined that $x(t)$ is approximately continuous in P-measure at almost all t-points.

Let $\Phi(t)$ be the function whose value for each t is the ω-function $x_t(\omega)$. Then from Theorems 1 and 2 we know that there is a measurable process if and only if $\Phi(t)$ is approximately continuous almost everywhere, that is, if and only if for almost every t there is an h-set of density 1 at $h = 0$ such that

$$\int |\arctan x_{t+h}(\omega) - \arctan x_t(\omega)| \, dP \to 0$$

as h approaches zero on that h-set. This condition is equivalent to the condition that for every $\epsilon > 0$

$$P(|\arctan x_{t+h}(\omega) - \arctan x_t(\omega)| > \epsilon) \to 0$$

and hence to the condition that for every $\epsilon > 0$

* Kolmogoroff has not yet published his proof of this theorem; it is stated in a footnote on page 757 of [4].

$$P(|x(t+h)-x(t)|>\epsilon)\to 0$$

so that the approximate continuity almost everywhere of $\Phi(t)$ is equivalent to this condition of Kolmogoroff; hence this condition is necessary and sufficient for the existence of a measurable process.

3. **Measurable stochastic processes and the conditional probability functions.** In this section we shall define the conditional probability functions of a stochastic process. We shall then obtain necessary and sufficient conditions upon them that Ω contain a measurable stochastic process. The definition which we use of these functions is due to Kolmogoroff [5, pp. 41–43].*

Let M be a P-measurable set and let $t_1, \cdots, t_n$ be a set of real numbers. Let N be a P-measurable set depending only upon $x_{t_1}, \cdots, x_{t_n}$† and consider the set-function $P(MN)$. As N ranges over the P-measurable sets which depend only upon $x_{t_1}, \cdots, x_{t_n}$, this is a completely additive nonnegative set-function which vanishes whenever $P(N)=0$. Hence there is a nonnegative P-measurable function which depends only upon $x_{t_1}, \cdots, x_{t_n}$ and which we shall denote by

$$P(x_{t_1}(\omega), \cdots, x_{t_n}(\omega); M)$$

such that

$$P(M\cdot N)=\int_N P(x_{t_1}(\omega), \cdots, x_{t_n}(\omega); M)dP.\ddagger$$

A discussion of the properties of these conditional probabilities is to be found in [5].

Before stating our next theorem we shall establish the following notation. If E is a set of real numbers, then we shall denote by E_t the ω-set, $_\omega[x_t(\omega)\ \epsilon\ E]$. If I is an open interval of real numbers, then the set of inner points of the complement of I is called the *open complement* of I and is denoted by J.

THEOREM 4. *If t is a real number and if E is a set (of real numbers) with zero as a limit point, then the following assertions are equivalent:*

(1) *For every $\epsilon>0$, $P(|x(t+h)-x(t)|>\epsilon)\to 0$ $(h\to 0, h\ \epsilon\ E)$.*

---

* In this connection, also see [2, 3].

† To say that N depends only upon $x_{t_1}, \cdots, x_{t_n}$ means that if $x(t)$ and $x'(t)$ are any two points of Ω and if $x(t_i)=x'(t_i)$ $(i=1, \cdots, n)$, then either both or neither of them belongs to N. To say that an ω-function depends only upon $x_{t_1}, \cdots, x_{t_n}$ means similarly that if $x(t)$ and $x'(t)$ are any two elements of Ω and if $x(t_i)=x'(t_i)$ $(i=1, \cdots, n)$, then the function has the same value at these two points.

‡ The existence of such a function is a consequence of the Radon-Nikodym theorem. (See [8, p. 36].) That the function depends (in this case) only upon $x_{t_1}, \cdots, x_{t_n}$ is readily verified. This function is of course uniquely defined only up to a set of measure zero.

(2) *For every open interval (of real numbers) I with open complement J*

$$P(I_t \cdot J_{t+h}) + P(J_t \cdot I_{t+h}) \to 0,^* \qquad h \to 0,\ h \in E.$$

Proof. We shall show first that (1) implies (2). Let I be any open interval with open complement J and let $\{I^n\}$ be the sequence of intervals defined by I^n = points at distance greater than $1/n$ from J so that the intervals I^n expand to I as $n \to \infty$. Then for $\omega \in I_t^n \cdot J_{t+h}$ we have $|x_t(\omega) - x_{t+h}(\omega)| > 1/n$ so that (1) implies, for fixed n,

$$P(I_t^n \cdot J_{t+h}) \to 0, \qquad h \to 0,\ h \in E. \tag{a}$$

Because the sets I^n expand to I we have, uniformly in h,

$$P(I_t^n \cdot J_{t+h}) \to P(I_t \cdot J_{t+h}). \tag{b}$$

Then (a) and (b) together imply that

$$P(I_t \cdot J_{t+h}) \to 0, \qquad h \to 0,\ h \in E.$$

In the same way it can be shown that $P(J_t \cdot I_{t+h}) \to 0$, $h \to 0$, $h \in E$, so that (2) must hold.

To show that (2) implies (1) we shall show that (with $\epsilon > 0$ arbitrarily preassigned) for any $\delta > 0$ there is an $\eta > 0$ such that

$$P(|x(t+h) - x(t)| > \epsilon) < \delta \quad \text{for} \quad |h| < \eta, \qquad h \in E.$$

Given $\delta > 0$, first choose M so large that $P(|x(t)| \geqq M) < \delta/2$. Then cover the interval $(-M, M)$ with a finite number of open intervals each of length less than ϵ. If p is the number of intervals in this covering, then we can find a positive number η so small that for every interval I of the covering, with open complement J, we have

$$P(I_t \cdot J_{t+h}) < \delta/2p, \quad \text{for} \quad |h| < \eta, \qquad h \in E,$$

and because of

$$_\omega[|x(t+h) - x(t)| > \epsilon] \subset \sum I_t \cdot J_{t+h} + {}_\omega[|x(t)| \geqq M]$$

we have

$$P(|x(t+h) - x(t)| > \epsilon) < \sum \delta/2p + \delta/2 = \delta \quad \text{for} \quad |h| < \eta, \qquad h \in E.$$

For our next theorem we shall have to consider for each interval I, with open complement J, the set $K = J + I$. We shall call K the set *corresponding to* I. With this understood we can state

* In case the end points a and b of I are such that the two ω-sets ${}_\omega[x_t(\omega) = a]$ and ${}_\omega[x_t(\omega) = b]$ both have P-measure zero, this condition can be replaced by the condition $P(I_t \cdot CI_{t+h}) + P(I_{t+h} \cdot CI_t) \to 0$.

THEOREM 5. *If t is a real number and E is a set (of real numbers) with zero as a limit point, then the following assertions are equivalent:*

(1) *For every open interval (of real numbers) I with open complement J,*

$$P(I_t \cdot J_{t+h}) + P(J_t \cdot I_{t+h}) \to 0, \qquad h \to 0,\ h \in E.$$

(2) *For every open interval I and corresponding set K, we have*

$$P(K_t \cdot {}_\omega[\,|\,P(x_t(\omega); I_{t+h}) - \varphi_I(x_t)\,| > \epsilon]) \to 0, \qquad h \to 0,\ h \in E,$$

*for every $\epsilon > 0$.**

Proof. To see that (1) implies (2) we note that $P(J_t \cdot I_{t+h}) \to 0$ says that

$$\int_{J_t} P(x_t(\omega); I_{t+h})\,dP \to 0;$$

and since this integrand is nonnegative, this implies that

$$\text{(a)} \qquad P(J_t \cdot {}_\omega[P(x_t(\omega); I_{t+h}) > \epsilon]) \to 0, \qquad h \to 0,\ h \in E.$$

From the fact that $P(I_t \cdot J_{t+h}) \to 0$ ($h \to 0$, $h \in E$) we can show similarly that

$$\text{(b)} \qquad P(I_t \cdot {}_\omega[\,|\,P(x_t(\omega); I_{t+h}) - 1\,| > \epsilon]) \to 0, \qquad h \to 0,\ h \in E.$$

Statements (a) and (b) together imply (2).

(2) says that $P(x_t(\omega); I_{t+h})$ converges, on K_t, in P-measure to $\varphi_I(x_t)$ as h approaches zero on E; hence

$$\int_{J_t} P(x_t(\omega); I_{t+h})\,dP \to \int_{J_t} \varphi_I(x_t)\,dP, \qquad h \to 0,\ h \in E,$$

that is, $P(J_t \cdot I_{t+h}) \to 0$, $h \to 0$, $h \in E$. It follows similarly that $P(I_t \cdot J_{t+h}) \to 0$ and hence (2) implies (1).

Combining Theorems 3, 4, and 5 we can state

THEOREM 6. *Each of the following conditions is necessary and sufficient for Ω to contain a measurable stochastic process:*

(1) *For almost every t there is an h-set of density 1 at $h = 0$ such that for every open interval of real numbers I with open complement J*

$$P(I_t \cdot J_{t+h}) + P(J_t \cdot I_{t+h}) \to 0$$

as h approaches zero on that h-set.

(2) *For almost every t there is an h-set of density 1 at $h = 0$ such that for every*

* In case the end points a and b of I are such that the ω-sets ${}_\omega[x_t(\omega) = a]$ and ${}_\omega[x_t(\omega) = b]$ both have P-measure zero, this condition can be replaced by the condition $P(|P(x_t(\omega); I_{t+h}) - \varphi_I(x_t)| > \epsilon) \to 0$. The symbol $\varphi_I(x_t)$ denotes the characteristic function of the ω-set I_t.

open interval of real numbers I, and corresponding set K, we have for every $\epsilon>0$

$$P(K_{t\cdot\omega}[\,|\,P(x_t(\omega);\,I_{t+h})-\varphi_I(x_t)\,|>\epsilon])\to 0$$

as h approaches zero on that h-set.

DEFINITION. *Suppose that for every open interval I of real numbers there is a function $\varphi(t,h)$ with the property that for almost all t there is an h-set of density 1 at $h=0$ such that $\varphi(t,h)\to 0$ as h approaches zero on that h-set, and such that for every real number s*

$$\int |\,P(x_s(\omega);\,I_{t+h})-P(x_s(\omega);\,I_t)\,|\,dP<\varphi(t,h).$$

Then we shall say that the conditional probabilities are uniformly dominated.

THEOREM 7. *Suppose P-measure so defined that for every pair of real numbers a and t the ω-set for which $x_t(\omega)=a$ has P-measure zero. Then Ω contains a measurable stochastic process if and only if the conditional probabilities are uniformly dominated.*

Proof. We shall prove this theorem by showing that under the hypothesis of this theorem the conditional probabilities are uniformly dominated if and only if condition (2) of Theorem 6 is satisfied. Under the hypothesis made here the open complement of I may be replaced in condition (2) of Theorem 6 by the complement of I. Suppose now that the conditional probabilities are uniformly dominated; then

$$\begin{aligned}P(CI_t\cdot I_{t+h}) &= P(CI_t\cdot I_{t+h})-P(CI_t\cdot I_t)\\ &=\int_{CI_t}\{P(x_t(\omega);\,I_{t+h})-P(x_t(\omega);\,I_t)\}dP\\ &\leqq\int_\Omega |\,P(x_t(\omega);\,I_{t+h})-P(x_t(\omega);\,I_t)\,|\,dP\\ &<\varphi(t,h).\end{aligned}$$

It can be shown similarly that $P(I_{t+h}\cdot CI_t)<\varphi(t,h)$. Hence condition (2) of Theorem 6 is satisfied.

Because we have, for any P-measurable set M,

$$|\,P(M\cdot I_t)-P(M\cdot I_{t+h})\,|\leqq P(I_t\cdot CI_{t+h})+P(I_{t+h}\cdot CI_t),$$

we have, for any P-measurable set F_s, depending only upon x_s

$$\left|\int_{F_s}P(x_s(\omega);\,I_t)dP-\int_{F_s}P(x_s(\omega);\,I_{t+h})dP\right|\leqq P(I_t\cdot CI_{t+h})+P(I_{t+h}\cdot CI_t).$$

Since this last inequality is independent of the set F_s, we can conclude that

$$\int \left| P(x_s(\omega); I_t) - P(x_s(\omega); I_{t+h}) \right| dP \leqq 2\{P(I_t \cdot CI_{t+h}) + P(I_{t+h} \cdot CI_t)\},$$

and from this inequality we see that if condition (2) of Theorem 6 holds, then the conditional probabilities must be uniformly dominated.

4. **A theorem of W. Doeblin.** We have been concerned so far throughout this paper with the existence of a measurable process, but we have not inquired into any of the properties of that process. In this section we shall prove a theorem, due essentially to W. Doeblin, concerning the existence of a measurable process whose elements satisfy a certain regularity condition.

We shall say that Ω is a Markoff process if whenever $t_1 < t_2 < \cdots < t_n$ are real numbers and M is a P-measurable set depending only upon $\{x_\alpha\}$, with all the α's greater than or equal to t_n, then

$$P(x_{t_1}(\omega), \cdots, x_{t_n}(\omega); M) = P(x_{t_n}(\omega); M).$$

If it is possible to define the conditional probability functions so that* for every $\epsilon > 0$ we have

$$P(x_t(\omega); \left| x_t(\omega) - x_{t+h}(\omega) \right| < \epsilon) \to 1$$

uniformly in ω and t, then we shall say that the conditional probabilities satisfy the *Doeblin condition*. In these terms we state the theorem of Doeblin:

THEOREM 8. *If Ω is a Markoff process whose conditional probabilities satisfy the Doeblin condition, then there is a measurable stochastic process whose elements $x(t)$ are continuous almost everywhere on the t axis, with probability* 1, *and $x(t)$ is continuous at $t = t_0$ with probability* 1, *for every value of t_0.*†

Proof. Because of Theorem 2.5 (i) of [2] it is sufficient to show that $t_n \to t$ implies $P(\lim_{n\to\infty} x(t_n) = x(t)) = 1$.‡ We shall show separately that $t_n \downarrow t$ implies $P(x(t_n) \to x(t)_{n\to\infty}) = 1$ and that $t_n \uparrow t$ implies $P(x(t_n) \to x(t)_{n\to\infty}) = 1$. From these facts it follows readily that $t_n \to t$ implies $P(x(t_n) \to x(t)_{n\to\infty}) = 1$.

First suppose $t_n \downarrow t$. Then define the ω-sets

$$M_n = {}_\omega[\text{for some } k \geqq n\colon \left| x(t_k) - x(t) \right| > \epsilon],$$

$$M_{nj} = {}_\omega[\text{for some } k \text{ among } n, n+1, \cdots, n+j\colon \left| x(t_k) - x(t) \right| > \epsilon],$$

$$M_{nji} = {}_\omega[\left| x(t_{n+i}) - x(t) \right| > \epsilon] \cdot \prod_{j \leqq l \leqq i+1} {}_\omega[\left| x(t_{n+l}) - x(t) \right| \leqq \epsilon].$$

* Because the conditional probabilities are only defined uniquely up to a set of measure zero, there will be, in general, infinitely many choices for them; if for any one of those choices this condition is fulfilled, then the conditional probabilities will be said to satisfy the Doeblin condition.

† See [1, pp. 49–53].

‡ Because we shall show this for *every* value of t it will follow (from the proof of Theorem 2.5 (i) of [2] that $x(t)$ is continuous at $t = t_0$ with probability 1, for *every* value of t_0.

We observe that $M_{nj}=\sum_{i=0}^{j}M_{nji}$, $M_{nji}\cdot M_{njk}=0$ for $i\neq k$, $M_{nj}\subset M_{n,j+1}$, $M_n=\sum_{j=0}^{\infty}M_{nj}$. It is easily seen that

$$\omega[\,|\,x(t_n)-x(t)\,|\,>\epsilon/2]\supset\sum_{i=0}^{j}M_{nji}\cdot\omega[\,|\,x(t_{n+i})-x(t_n)\,|\,<\epsilon/2]$$

and hence that

$$(\alpha)\qquad\begin{aligned}P(\,|\,x(t_n)-x(t)\,|\,>\epsilon/2)&\geqq P\Big(\sum_{i=0}^{j}M_{nji}\cdot\omega[\,|\,x(t_{n+i})-x(t_n)\,|\,<\epsilon/2]\Big)\\&=\sum_{i=0}^{j}P(M_{nji}\cdot\omega[\,|\,x(t_{n+i})-x(t_n)\,|\,<\epsilon/2]).\end{aligned}$$

Now because M_{nji} depends only upon x_t, $x_{t_{n+j}}$, $\cdots$, $x_{t_{n+i}}$ and because the process is a Markoff process, we have

$$\begin{aligned}&P(M_{nji}\cdot\omega[\,|\,x(t_{n+i})-x(t_n)\,|\,<\epsilon/2])\\&\qquad=\int_{M_{nji}}P(x_{t_{n+i}}(\omega);\ |\,x(t_{n+i})-x(t_n)\,|\,<\epsilon/2)dP.\end{aligned}$$

If we choose n sufficiently large, the function

$$P(x_{t_{n+i}}(\omega);\ |\,x(t_{n+i})-x(t_n)\,|\,<\epsilon/2)$$

will be greater than $1/2$ uniformly in ω and i, since the conditional probabilities satisfy the Doeblin condition. Hence for n sufficiently large we have, unifomly in i,

$$P(M_{nji}\cdot\omega[\,|\,x(t_{n+i})-x(t_n)\,|\,<\epsilon/2])>(1/2)P(M_{nji}),$$

and using this fact in the inequality (α) we have

$$P(\,|\,x(t_n)-x(t)\,|\,>\epsilon/2)\geqq\sum_{i=0}^{j}(1/2)P(M_{nji})=(1/2)P(M_{nj}).$$

Hence

$$P(M_{nj})\leqq 2P(\,|\,x(t_n)-x(t)\,|\,>\epsilon/2)$$

for n sufficiently large, uniformly in j. Since the sets M_{nj} expand (as $j\to\infty$) to M_n, we have

$$(\beta)\qquad P(M_n)\leqq 2P(\,|\,x(t_n)-x(t)\,|\,>\epsilon/2).$$

Because the Doeblin condition implies that for every open interval I with corresponding set K

$$P(K_t\cdot\omega[\,|\,P(x_t(\omega);I_{t_n})-\varphi_I(x_t)\,|\,>\epsilon])\to 0\qquad(n\to\infty),$$

we see by Theorems 4 and 5 that $P(|x(t_n)-x(t)|>\epsilon/2)\to 0$ $(n\to\infty)$ so that (β) implies $P(M_n)\to 0$ $(n\to\infty)$. This implies that $P(\lim_{n\to\infty} x(t_n)=x(t))=1$.

We shall now show that $t_n\uparrow t$ implies $P(x(t_n)\to x(t)_{n\to\infty})=1$. For each positive integer n we define the sequence of sets

$$A_n^{n+1}=\omega[\,|\,x(t_{n+1})-x(t_n)\,|\,>\epsilon],$$

$$\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots,$$

$$A_n^{n+j}=\omega[\,|\,x(t_{n+j})-x(t_n)\,|\,>\epsilon]\cdot\prod_{i=1}^{j-1}\omega[\,|\,x(t_{n+i})-x(t_n)\,|\,\leqq\epsilon],$$

$$\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots,$$

and we define the set A_n by

$$A_n=\omega[\text{for some } k>n\colon\ |\,x(t_k)-x(t_n)\,|\,>\epsilon].$$

We note that A_n^{n+j} depends only upon $x_{t_n},\cdots,x_{t_{n+j}}$. Also we have $A_n=\sum_{j=1}^{\infty}A_n^{n+j}$. It is easily seen that

$$\omega[\,|\,x(t_n)-x(t)\,|\,>\epsilon/2]\supset\sum_{j=1}^{\infty}A_n^{n+j}\cdot\omega[\,|\,x(t_{n+j})-x(t)\,|\,<\epsilon/2]$$

and hence that

$$(\gamma)\qquad\begin{aligned}P(\,|\,x(t_n)-x(t)\,|\,>\epsilon/2)&\geqq P\Big(\sum_{j=1}^{\infty}A_n^{n+j}\cdot\omega[\,|\,x(t_{n+j})-x(t)\,|\,<\epsilon/2]\Big)\\&=\sum_{j=1}^{\infty}P(A_n^{n+j}\cdot\omega[\,|\,x(t_{n+j})-x(t)\,|\,<\epsilon/2]).\end{aligned}$$

Now because A_n^{n+j} depends only upon $x_{t_n},\cdots,x_{t_{n+j}}$ we have

$$P(A_n^{n+j}\cdot\omega[\,|\,x(t_{n+j})-x(t)\,|\,<\epsilon/2])=\int_{A_n^{n+j}}P(x_{t_{n+j}}(\omega);\ |\,x(t_{n+j})-x(t)\,|\,<\epsilon/2)dP.$$

As above we see that for n sufficiently large this integrand is greater than $1/2$, uniformly in j, and hence for n sufficiently large

$$P(A_n^{n+j}\cdot\omega[\,|\,x(t_{n+j})-x(t)\,|\,<\epsilon/2])>(1/2)P(A_n^{n+j}).$$

This implies when used in (γ) that

$$P(\,|\,x(t_n)-x(t)\,|\,>\epsilon/2)\geqq\sum_{j=1}^{\infty}(1/2)P(A_n^{n+j})=(1/2)P(A_n)$$

for n sufficiently large. The Doeblin condition implies, by virtue of Theorems 4 and 5, that the expression on the left tends to zero as $n\to\infty$ and hence that $P(A_n)\to 0$ as $n\to\infty$. This means that the sequence $\{x(t_n)\}$ satisfies the Cauchy condition, with probability 1, and hence converges, with probability 1, as $n\to\infty$. That it converges to $x(t)$ with probability 1 then follows from the fact that the left side of the above expression tends to 0 as n becomes infinite.

Bibliography

1. W. Doeblin, *Sur les propriétés asymptotiques de mouvement regis par certains types des chaînes simples*, Bulletin Mathématique de la Société Roumaine des Sciences, vol. 39, no. 1, pp. 57–115; no. 2, pp. 3–61.

2. J. L. Doob, *Stochastic processes depending upon a continuous parameter*, these Transactions, vol. 42 (1937), pp. 107–113.

3. ———, *Stochastic processes with an integral-valued parameter*, these Transactions, vol. 44 (1938), pp. 87–150.

4. ———, *One-parameter families of transformations*, Duke Mathematical Journal, vol. 4 (1938), pp. 752–774.

5. A. Kolmogoroff, *Grundbegriffe der Wahrscheinlichkeitsrechnung*, Ergebnisse der Mathematik und ihrer Grenzgebiete, vol. 2, no. 3.

6. S. Saks, *Theory of the Integral*, New York, 1939.

7. E. Slutsky, *Qualche proposizione relativa alla teoria delle funzioni aleatorie*, Giornale dell' Istituto degli Attuari, vol. 8 (1937), pp. 182–199.

University of Illinois,
Urbana, Ill.

CONTRIBUTIONS TO THE THEORY OF HERMITIAN SERIES

II. THE REPRESENTATION PROBLEM*

BY

EINAR HILLE

1. **Introduction.** In the first note of this series [2]† the author laid a broad foundation for the theory of Hermitian series in the complex domain. A number of questions encountered during this investigation were merely mentioned and detailed discussion had to be postponed till later communications. The *representation* or *expansion problem* was such a question, that is, *the problem of finding necessary and sufficient conditions in order that an analytic function shall be representable by an Hermitian series for complex values of the variable.* This problem is solved in the present note.

Let $\boldsymbol{H}_n(z)$ denote the nth normalized orthogonal function of Hermite

$$\boldsymbol{H}_n(z) = [\pi^{1/2}2^n n!]^{-1/2}(-1)^n e^{z^2/2}\frac{d^n}{dz^n}[e^{-z^2}]. \tag{1.1}$$

If $f(x)$ is a measurable function such that $x^n \exp[-x^2/2]\, f(x) \in L_1(-\infty, \infty)$ for $n=0, 1, 2, \cdots$, then $f(x)$ has an associated Fourier-Hermite series

$$f(x) \sim \sum_{n=0}^{\infty} f_n \boldsymbol{H}_n(x) \tag{1.2}$$

where

$$f_n = \int_{-\infty}^{\infty} f(t)\boldsymbol{H}_n(t)dt.$$

This series is ordinarily not convergent, but if we assume for instance that $\exp[-\alpha x^2] f(x) \in L_1(-\infty, \infty)$ for some $\alpha < 1/2$, then the series (1.2) can be summed to the sum $f(x)$ for almost all x by a generalization of the Abel method of summation.‡ Under this assumption the series

$$f(x; s) = \sum_{n=0}^{\infty} f_n \boldsymbol{H}_n(x)s^n$$

converges for

---

* Presented to the Society, October 28, 1939; received by the editors November 6, 1939.

† Numbers in square brackets refer to the references at the end of this paper.

‡ See E. Hille [1, pp. 448–450, 453].

$$|s| < \left(\frac{1-2\alpha}{1+2\alpha}\right)^{1/2} \equiv r_\alpha.$$

Further $f(x; s)$ is a holomorphic function of s in the interior of the ellipse in the complex s-plane having its vertices at the points $s = \pm 1, \pm ir_\alpha$. As $s \to +1$ along the real axis, $f(x; s) \to f(x)$ for almost all values of x. It follows in particular that if the series (1.2) converges in a set E of positive measure, its sum equals $f(x)$ almost everywhere in E provided $\exp[-\alpha x^2] f(x) \in L_1(-\infty, \infty)$ for some $\alpha < 1/2$.

It may happen that the series (1.2) converges also for complex values of the variable. If so, it is known that the domain of absolute convergence is a strip S_τ: $-\tau < y < \tau$, $z = x + iy$, where

$$\tau = -\limsup_{n\to\infty} (2n+1)^{-1/2} \log |f_n|. \tag{1.3}$$

For a proof, see E. Hille [2, chap. 2] and G. Szegö [7, p. 246]. The series may converge on the lines of convergence $y = \pm\tau$, but it diverges outside of S_τ. In this case the series

$$\sum_{n=0}^{\infty} f_n \boldsymbol{H}_n(z) \equiv f(z) \tag{1.4}$$

defines an analytic function which is holomorphic in S_τ.

Such a function $f(z)$ is of course characterized by the fact that its Fourier-Hermite coefficients satisfy the inequality

$$|f_n| \leqq M(\epsilon) \exp[-(\tau - \epsilon)(2n+1)^{1/2}] \tag{1.5}$$

for every positive ϵ, but a direct characterization of $f(z)$ in terms of function-theoretical properties would seem desirable. A complete solution of this problem is given by

THEOREM 1. *Let $f(z)$ be an analytic function. A necessary and sufficient condition in order that the Fourier-Hermite series*

$$\sum_{n=0}^{\infty} f_n \boldsymbol{H}_n(z), \qquad f_n = \int_{-\infty}^{\infty} f(t) \boldsymbol{H}_n(t)\, dt, \tag{1.6}$$

shall exist and converge to the sum $f(z)$ in the strip S_τ: $-\tau < y < \tau$, is that $f(z)$ is holomorphic in S_τ and that to every given β, $0 \leqq \beta < \tau$, there exists a finite positive $B(\beta)$ such that

$$|f(x+iy)| \leqq B(\beta) \exp[-|x|(\beta^2 - y^2)^{1/2}] \tag{1.7}$$

for $-\infty < x < \infty$, $-\beta \leqq y \leqq \beta$.

The proof that this condition is necessary will be given in §2. The sufficiency proof occurs in §5 and is preceded by material relating to the asymptotic behavior of $H_n(z)$ for large values of n based on the investigations of R. E. Langer [3]. An application to functions meromorphic in an upper or a lower half-plane is given in §6.

As far as I know there have only been two previous contributions to the representation problem for Hermitian series, those of G. N. Watson [9, pp. 417–421] and O. Volk [8]. Their equivalent conditions require that $f(z)$ shall be holomorphic in $-\tau \leqq y \leqq \tau$ and that the function $g(z) \equiv z^{-1} \exp(z^2/2) f(z)$ shall have the properties (i) $g(x+iy) \to 0$ as $|x| \to \infty$, $-\tau \leqq y \leqq \tau$, and (ii) $g(x \pm i\tau) \in L_1(-\infty, \infty)$. It is clear that these conditions while sufficient are far from necessary.

2. **Proof of necessity.** The reader will be supposed to have some knowledge of the properties of Hermitian polynomials and of solutions of the Weber-Hermite differential equations. For convenient summaries of the former theory see E. Hille [1] and G. Szegö [7], for the latter see Whittaker-Watson [12].

The only property which will be used in the present paragraph is the generating function*

$$\sum_{n=0}^{\infty} H_n(u) H_n(v) s^n = \pi^{-1/2}(1-s^2)^{-1/2} \exp\left\{-\frac{(1+s^2)(u^2+v^2)-4suv}{2(1-s^2)}\right\}. \tag{2.1}$$

The series converges for arbitrary complex values of u and v when $|s|<1$. Putting $u=z=x+iy$, $v=\bar{z}=x-iy$, we get

$$\sum_{n=0}^{\infty} |H_n(z)|^2 s^n = \pi^{-1/2}(1-s^2)^{-1/2} \exp\left\{-\frac{1-s}{1+s}x^2 + \frac{1+s}{1-s}y^2\right\}. \tag{2.2}$$

We shall also need information concerning integrals of the form

$$J(p, q, \nu) = \int_0^{\infty} \exp\{-pt - q/t\}\, t^{-\nu} dt,$$

where p and q are positive and $\nu = 3/2$, 2 or $5/2$. Now

$$J(p, q, 3/2) = \pi^{1/2} q^{-1/2} \exp\left[-2(pq)^{1/2}\right], \tag{2.3}$$

as is easily verified. Since $J(p, q, 5/2) = -J_q'(p, q, 3/2)$,

$$J(p, q, 5/2) = \pi^{1/2} q^{-3/2}\left[1/2 + (pq)^{1/2}\right] \exp\left[-2(pq)^{1/2}\right]. \tag{2.4}$$

* See, for instance, E. Hille [1, pp. 439–440].

Finally, by Schwarz' inequality

$$J(p, q, 2) \leqq \pi^{1/2}q^{-1}[1/2 + (pq)^{1/2}]^{1/2} \exp\,[-2(pq)^{1/2}]. \tag{2.5}$$

Suppose now that the series (1.6) exists and converges to $f(z)$ in the strip $-\tau<y<\tau$. It follows immediately that $f(z)$ is holomorphic in the strip and it remains merely to estimate $f(z)$. Let β be given, $0<\beta<\tau$, and put $\alpha=(\beta+\tau)/2$. In view of (1.5), the series

$$\sum_{n=0}^{\infty} |f_n|^2 \exp(2\alpha N) \equiv A^2(\beta) \tag{2.6}$$

is convergent. Here and in the following we shall write

$$N = (2n + 1)^{1/2}. \tag{2.7}$$

Let us put

$$H(x, y; \alpha) = \sum_{n=0}^{\infty} |H_n(x + iy)|^2 \exp(-2\alpha N). \tag{2.8}$$

Cauchy's inequality then gives

$$|f(x + iy)| \leqq A(\beta)[H(x, y; \alpha)]^{1/2}. \tag{2.9}$$

Let us first obtain $H(x, y; \alpha)$ in closed form. Forming

$$\sum_{n=0}^{\infty} |H_n(z)|^2 J(2n + 1, \alpha^2, 3/2)$$

and using (2.2) and (2.3) we get after some simplifications

$$H(x, y; \alpha) = \frac{\alpha}{\pi}\int_0^{\infty} (1 - e^{-4t})^{-1/2} \exp\{-\alpha^2/t - t - x^2 \tanh t + y^2 \coth t\} t^{-3/2}dt, \tag{2.10}$$

where $\tanh t$ denotes the hyperbolic tangent of t.

Elementary considerations show that

$$\coth t \leqq (1 + t)/t, \qquad -\tanh t \leqq -t/(1 + t), \qquad t > 0.$$

Substituting these dominants for the hyperbolic functions in (2.10) we obtain

$$\begin{aligned}(\pi/\alpha)H(x, y; \alpha) &\leqq H_0(x, y; \alpha) \\ &\equiv e^{y^2}\int_0^{\infty} (1 - e^{-4t})^{-1/2} \exp\left\{-\frac{\alpha^2 - y^2}{t} - t - \frac{x^2 t}{1 + t}\right\} t^{-3/2}dt \\ &\equiv e^{y^2}H_1(x, y; \alpha),\end{aligned} \tag{2.11}$$

and it remains to estimate $H_1(x, y; \alpha)$.

Let $\xi = 4\alpha|x|^{-1}$ and put

$$H_1 = H_{11} + H_{12}, \quad H_{11} = \int_0^{\xi} \cdots, \quad H_{12} = \int_{\xi}^{\infty} \cdots \tag{2.12}$$

with obvious notation.

We note that $t/(1+t)$ is an increasing function. Hence its least value in the interval (ξ, ∞) is taken on at $t=\xi$. The inequality $1+u \leqq e^u$ gives

$$(1+u)^{-1} \geqq e^{-u}, \qquad (1-e^{-u})^{-1} \leqq 1/u + 1$$

and finally

$$(1-e^{-4t})^{-1/2} \leqq (1/(4t)+1)^{1/2} \leqq t^{-1/2}/2 + 1.$$

Hence

$$H_{12} \leqq \exp\left[-\frac{4\alpha x^2}{4\alpha + |x|}\right]\{J(1, \alpha^2 - y^2, 3/2) + (1/2)J(1, \alpha^2 - y^2, 2)\}.$$

Supposing $|x| \geqq 4\alpha$, which is no essential restriction, and using (2.3) and (2.5), we get

$$H_{12} \leqq \pi^{1/2} \exp\left[-2\alpha|x| - 2(\alpha^2 - y^2)^{1/2}\right](\alpha^2 - y^2)^{-1} \cdot \{(1/2)[1/2 + (\alpha^2 - y^2)^{1/2}]^{1/2} + (\alpha^2 - y^2)^{1/2}\}.$$

If $-\beta \leqq y \leqq \beta$, the right-hand side does not exceed

$$\frac{10\beta + 7}{(\tau + 3\beta)(\tau - \beta)} \exp\left[-2\alpha|x|\right],$$

that is,

$$H_{12}(x, y; \alpha) \leqq B_1(\beta) \exp\left[-2\alpha|x|\right] \tag{2.13}$$

for $|x| \geqq 4\alpha$, $-\beta \leqq y \leqq \beta$, and by a suitable modification of $B_1(\beta)$ we can of course gain that the inequality is true for all x. Here and in the following $B(\beta)$, with or without subscripts, denotes a monotone increasing function of β, defined and continuous for $0 \leqq \beta < \tau$, but conceivably tending to infinity as $\beta \to \tau$.

For $t > 0$ we have

$$-t/(1+t) = -t + t^2/(1+t) < -t + t^2.$$

Further, for $0 < t \leqq \xi \leqq 1$, that is, for $|x| \geqq 4\alpha$,

$$(1-e^{-4t})^{-1/2} \leqq (1-e^{-4})^{-1/2} t^{-1/2} < 2t^{-1/2}.$$

Hence

$$H_{11} < 2e^{t^2x^2}\int_0^{\xi} \exp\left\{-(\alpha^2 - y^2)/t - x^2t\right\}t^{-2}dt$$
$$< 2e^{16\alpha^2}J(x^2, \alpha^2 - y^2, 2)$$
$$< 2\pi^{1/2}e^{16r^2}(\alpha^2 - y^2)^{-1}[1/2 + |x|(\alpha^2 - y^2)^{1/2}]^{1/2}\exp[-2|x|(\alpha^2 - y^2)^{1/2}]$$
$$< 4e^{16r^2}(\alpha^2 - y^2)^{-1}[1 + |x|^{1/2}(\alpha^2 - y^2)^{1/4}]\exp[-2|x|(\alpha^2 - y^2)^{1/2}].$$

Suppose again that $|y| \leqq \beta$. Then $(\beta^2-y^2)^{1/2}<(\alpha^2-y^2)^{1/2}$ and the absolute maximum of

$$|x|^{1/2}(\alpha^2 - y^2)^{1/4}\exp\left\{-2|x|\left[(\alpha^2 - y^2)^{1/2} - (\beta^2 - y^2)^{1/2}\right]\right\}$$

for all values of x equals

$$\frac{1}{2}\left\{\frac{\alpha}{e(\alpha - \beta)}\right\}^{1/2}.$$

Consequently

$$H_{11} < 4e^{16r^2}(\alpha^2 - \beta^2)^{-1}[1 + \alpha^{1/2}(\alpha - \beta)^{-1/2}]\exp[-2|x|(\beta^2 - y^2)^{1/2}]$$

or

$$H_{11}(x, y; \alpha) < B_2(\beta)\exp[-2|x|(\beta^2 - y^2)^{1/2}] \tag{2.14}$$

for $-\beta \leqq y \leqq \beta$ and all values of x, since clearly the restriction $|x| \geqq 4\alpha$ can be removed.

Combining the inequalities in (2.9) to (2.14) we get the desired estimate

$$|f(x + iy)| < B(\beta)\exp[-|x|(\beta^2 - y^2)^{1/2}],$$

valid for every z in the strip $-\beta \leqq y \leqq \beta$ and for every $\beta < \tau$. This completes the proof of the necessity of condition (1.7).

3. **Some asymptotic formulas.** The sufficiency proof requires accurate information concerning the behavior of $H_n(z)$ for large values of n when the distance of z from the line segment $(-N, N)$ remains bounded. This question will be studied in this and the next following sections.

The Weber-Hermite differential equation

$$w'' + (2\kappa + 1 - z^2)w = 0, \tag{3.1}$$

where κ is an arbitrary real or complex parameter, has a solution $h_\kappa(z)$ which is uniquely characterized by the asymptotic property

$$\lim e^{z^2/2}(2z)^{-\kappa}h_\kappa(z) = 1 \tag{3.2}$$

when $|z| \to \infty$ and $|\arg z| < 3\pi/4$. Here the power has its principal determination when z and κ are real positive. In the notation of Whittaker and Watson

$$h_\kappa(z) = 2^{\kappa/2}D_\kappa(2^{1/2}z).$$

Since $H_n(z)$ satisfies (3.1) when $\kappa=n$ and

$$H_n(z) = (\pi^{1/2}2^n n!)^{-1/2}e^{-z^2/2}[(2z)^n + \cdots],$$

we conclude that

$$H_n(z) = (\pi^{1/2}2^n n!)^{-1/2}h_n(z). \tag{3.3}$$

It was shown by Whittaker [11] that $h_{-\kappa-1}(iz)$ and $h_{-\kappa-1}(-iz)$ are also solutions of (3.1) and Watson [9, p. 395] proved that

$$h_\kappa(z) = \pi^{-1/2}\Gamma(\kappa+1)2^\kappa\left\{e^{\kappa\pi i/2}h_{-\kappa-1}(iz) + e^{-\kappa\pi i/2}h_{-\kappa-1}(-iz)\right\}. \tag{3.4}$$

The asymptotic behavior of $h_{-\kappa-1}(-iz)$ was studied by Watson on two occasions. The result of his earlier paper [9, p. 416] is not good enough for our purposes, but that of his second paper [10, p. 142] more than meets our needs. Watson's analysis is based upon special representations of the functions involved by means of definite integrals to which he applies the method of steepest descent.

I prefer a different approach to the problem based upon the work of Langer [3] on the asymptotic behavior of solutions of certain classes of second order linear differential equations for large values of a parameter. This is a much more general approach and brings out the reason why different expansions are to be expected in different parts of the plane—the so-called phenomenon of Stokes. N. Schwid [6] has proved that Langer's method applies to the Weber-Hermite equation. Schwid does not discuss $h_{-\kappa-1}(\pm iz)$ explicitly, however, and in that part of the plane where we require information his results are not valid.* It is not necessary to give a detailed presentation of Langer's method. We shall merely state the results and refer to Langer and Schwid for all details.

In the following, κ shall be real positive. We write

$$\mathrm{K} = (2\kappa+1)^{1/2} \tag{3.5}$$

and†

$$\xi(z) = \int_{\mathrm{K}}^{z} [\mathrm{K}^2 - t^2]^{1/2}dt. \tag{3.6}$$

* His estimate of the order of the remainder term in formula (26), p. 353, is valid for bounded values of z only.

† Strictly speaking we should take $-\mathrm{K}$ as the lower limit of the integral when z is in the left half-plane. Since $\xi(-\mathrm{K})$ is a real quantity, the imaginary part of $\xi(z)$ is not affected by this change and the only place where it matters is in the remainder terms of formulas (3.7), (3.9), and (3.11) below. Here we should write $\xi-\xi(-\mathrm{K})$ instead of ξ in the left half-plane. The correction is of no importance, however, except in a small neighborhood of $z=-\mathrm{K}$ which is excluded from consideration anyway.

The Langer theory shows that for any value of $z \neq \pm K$ there are two linearly independent local solutions of (3.1) which for large values of κ are of the form

$$[\xi'(z)]^{-1/2} \exp [\pm i\xi(z)]\{1 + O(1/\xi) + O(1/\kappa)\} \tag{3.7}$$

and with the aid of which any solution can be expressed linearly. The coefficients will of course depend upon the determination chosen for the infinitely many-valued function $\xi(z)$. In the present case the situation is simplified by the fact that the imaginary part of $\xi(z)$ is only two-valued and the arcs $\Im[\xi(z)]=0$ are independent of the determination of $\xi(z)$. These arcs form a configuration symmetric with respect to the coordinate axes and divide the plane into four sectors S_j. The sector S_2, of main importance to us, is bounded by the line segment $(-K, K)$ and two transcendental curves Γ_1 and Γ_2 going from $\pm K$ to ∞. The angle of inclination of the tangent of Γ_1 with the positive real axis starts with the value $\pi/3$ at $z=K$ and tends to $\pi/4$ as $|z| \to \infty$.

Suppose the determination of $\xi(z)$ to be chosen at some point z_0 in S_2 and let a particular solution $w(z)$ of (3.1) be expressed linearly in terms of the two local solutions (3.7) in some neighborhood of $z=z_0$. Both solutions can be continued analytically in S_2, the continuations agreeing everywhere with the local fundamental system. It follows that the expression of $w(z)$ remains valid at all interior points of S_2. It is only when we cross the boundary of S_2 that the situation may change and a different substitution may have to be used to express $w(z)$ in terms of the local fundamental system. Thus to every sector S_j and to every given solution $w(z)$ there are corresponding coefficients C_{jk}, $j=1, 2, 3, 4$; $k=1, 2$, joining $w(z)$ with the fundamental system in S_j.

After these generalities let us choose $\xi(z)$ so that on the positive imaginary axis we have

$$\xi(iy) = -\pi K^2/4 + (i/2)\{y(K^2 + y^2)^{1/2} + K^2 \log [y/K + (1 + y^2/K^2)^{1/2}]\}, \tag{3.8}$$

where the square roots and the logarithm have their principal values. We can then take as the fundamental system in S_2 two solutions $w_{2,k}(z)$ which for large values of κ have the form

$$w_{2,k}(z) = \{1 - z^2/K^2\}^{-1/4} \exp [\pm i\xi(z)]\{1 + O(1/\xi) + O(1/\kappa)\}. \tag{3.9}$$

Here the plus sign goes with $k=1$ and the minus sign with $k=2$ and the fourth root is real positive on the positive imaginary axis. The parameter κ is supposed to be large and this will make $|\xi(z)|$ large if we omit small neighborhoods of the points $z=\pm K$ of radius $O(\kappa^{-1/6})$ inside of which (3.9) is not valid.*

* Compare the second footnote on page 86.

We have consequently

$$h_{-\kappa-1}(-iz) = A(\kappa)w_{2,1}(z) + B(\kappa)w_{2,2}(z) \tag{3.10}$$

in S_2. In order to determine $A(\kappa)$ and $B(\kappa)$ we use (3.2). Multiply both sides of (3.10) by $e^{-z^2/2}(-2iz)^{1+\kappa}$ and let $z\to\infty$, say along the positive imaginary axis. The left-hand side tends to the limit 1, whereas the right-hand side tends to a limit if and only if $B(\kappa)=0$. Setting $B(\kappa)=0$ and using (3.8) and (3.9), we can compute the value of $A(\kappa)$. The result is

$$\begin{aligned} h_{-\kappa-1}(-iz) = {} & 2^{-1/2}\mathrm{K}^{-\kappa-1}\exp\left[(1/4)\mathrm{K}^2(1+\pi i)\right] \\ & \cdot\left\{1-z^2/\mathrm{K}^2\right\}^{-1/4}\exp\left[i\xi(z)\right]\left\{1+O(1/\xi)+O(1/\kappa)\right\}, \end{aligned} \tag{3.11}$$

valid in S_2 except in neighborhoods of $z=\pm\mathrm{K}$ of radius $O(\kappa^{-1/6})$. This is the required basic asymptotic formula. It can be shown to be in agreement with Watson's formula quoted above.

4. **Estimates of the imaginary part of $\xi(z)$.** We shall need a couple of lemmas for the sufficiency proof.

LEMMA 1. *Let $\xi(x)$ be defined by* (3.6) *and let $-\mathrm{K}<x<\mathrm{K}$, $0\leqq y\leqq\mathrm{K}$. Then*

$$\begin{aligned} \Im[\xi(x+iy)] = {} & y(\mathrm{K}^2-x^2)^{1/2} + (1/6)y^3(\mathrm{K}^2-x^2)^{-1/2} \\ & + \rho\mathrm{K}^2y^3(\mathrm{K}^2-x^2)^{-3/2}, \end{aligned} \tag{4.1}$$

where $|\rho|\leqq 5/24$.

Proof. We have

$$\begin{aligned} \Im[\xi(x+iy)] &= \Im\left\{\int_x^{x+iy}(\mathrm{K}^2-t^2)^{1/2}dt\right\} \\ &= \Re\left\{\int_0^y\left[\mathrm{K}^2-x^2-2ixv+v^2\right]^{1/2}dv\right\}, \end{aligned}$$

where the value of the square root is real positive when $v=0$. But by elementary transformations

$$(a^2+b)^{1/2}-a-\frac{b}{2a} = -\frac{b^2}{4a^2}\left[(a^2+b)^{1/2}+a+\frac{b}{2a}\right]^{-1}.$$

Substituting $a^2=\mathrm{K}^2-x^2$, $b=-2ixv+v^2$, we find that the right-hand side becomes equal to P/Q where

$$\begin{aligned} P &= v^2(v-2ix)^2, \\ Q &= 4(\mathrm{K}^2-x^2)\left\{\left[\mathrm{K}^2-x^2-2ixv+v^2\right]^{1/2}+(\mathrm{K}^2-x^2)^{1/2}\right. \\ &\qquad \left.+(v^2-2ixv)(\mathrm{K}^2-x^2)^{-1/2}\right\}. \end{aligned}$$

Here $0 \leqq v \leqq y \leqq K$. The least value of $|Q|$ is obtained for $v=0$. On the other hand, $|v-2ix|^2 \leqq 5K^2$. Hence $|P/Q|$ does not exceed

$$(5/8)K^2(K^2 - x^2)^{-3/2}v^2,$$

the integral of which from $v=0$ to $v=y$ equals

$$(5/24)K^2(K^2 - x^2)^{-3/2}y^3.$$

This is the remainder term of formula (4.1). The other two terms are the result of substituting $a^2=K^2-x^2$, $b=-2ixv+v^2$ into $a+b/(2a)$, integrating with respect to v from 0 to y and finally taking the real part.

LEMMA 2. *For a fixed x, where $-K<x<K$, the maximum value of*

$$\Im[\xi(x+iy)] + |x|(\beta^2 - y^2)^{1/2}, \tag{4.2}$$

when $0 \leqq y \leqq \beta$, is not less than

$$\beta K + \frac{\beta^3}{6K^3}(K^2 - x^2) - \frac{5}{24}\frac{\beta^3}{K}. \tag{4.3}$$

In particular, this value is reached or exceeded on the ellipse

$$E(K, \beta): \quad \beta^2x^2 + K^2y^2 = \beta^2K^2. \tag{4.4}$$

Proof. By Lemma 1

$$\Im[\xi(x+iy)] \geqq y(K^2 - x^2)^{1/2} + (1/6)y^3(K^2 - x^2)^{-1/2} - (5/24)K^2y^3(K^2 - x^2)^{-3/2}.$$

The maximum value of

$$y(K^2 - x^2)^{1/2} + |x|(\beta^2 - y^2)^{1/2}$$

for a fixed x is reached when $y=(\beta/K)(K^2-x^2)^{1/2}$, that is, on the ellipse $E(K, \beta)$, and equals βK. Substitution of this value of y gives formula (4.3).

5. **The sufficiency proof.** After all these preparations we can attack the sufficiency proof. It is assumed that $f(z)$ is holomorphic in the strip S_τ: $-\tau<y<\tau$, and that to every β, $0<\beta<\tau$, there exists a finite positive $B(\beta)$ for which condition (1.7) holds.* This condition is evidently sufficient to ensure the existence of the Fourier-Hermite coefficients f_n defined by the second half of formula (1.6). Thus the Fourier-Hermite series of $f(z)$ has a sense, and condition (1.7) also ensures that this series is Abel summable to the sum $f(x)$ everywhere on the real axis. In order to complete the proof of the theorem it

* It is of course only values of β close to τ which are of interest, but it is no restriction to assume that $B(\beta)$ satisfies the conventions of §2 for all β, $0 \leqq \beta < \tau$. The same conventions can be imposed on all B-functions introduced below.

is then enough to prove that the series is convergent in the strip S_τ, and this will be proved if we are able to show the existence, for every β, $0<\beta<\tau$, of a finite positive $B^*(\beta)$ such that

$$|f_n| < B^*(\beta)e^{-\beta N}, \qquad N = (2n+1)^{1/2}, \tag{5.1}$$

for all n.

It is known that there exists a constant A such that†

$$|\boldsymbol{H}_n(t)| \leqq A \tag{5.2}$$

for all integers n and all real values of t. This inequality shows that

$$\left\{\int_{-\infty}^{-N+1} + \int_{N-1}^{\infty}\right\} |f(t)|\,|\boldsymbol{H}_n(t)|\,dt \leqq 2AB(\beta)\int_{N-1}^{\infty} e^{-\beta t}dt < B_1(\beta)e^{-\beta N}. \tag{5.3}$$

On the interval $(-N+1, N-1)$ we use formula (3.4) obtaining

$$\int_{-N+1}^{N-1} f(t)\boldsymbol{H}_n(t)dt = \pi^{-3/4}(2^n n!)^{1/2}\left\{i^n\int_{-N+1}^{N-1} f(t)h_{-n-1}(it)dt + (-i)^n\int_{-N+1}^{N+1} f(t)h_{-n-1}(-it)dt\right\}. \tag{5.4}$$

In these integrals we can choose the path of integration joining $-N+1$ with $N-1$ in an arbitrary manner provided it stays in the strip S_τ. Moreover, we can choose a different path for each of the integrals involved. Let $E(N, \beta)$ be the ellipse of Lemma 2 and denote the points of intersection of the ellipse with the lines $x=\pm(N-1)$ by P, Q, R, and S, where Q is in the first and P in the second quadrant.

The path of integration C_2 for the integral involving $h_{-n-1}(-it)$ shall consist of the line segments joining $-N+1$ with P and $N-1$ with Q connected by the arc PQ of the ellipse. The path of integration C_1 for the integral involving $h_{-n-1}(it)$ will be the image of C_2 in the real axis. Let us put

$$J_2 = \int_{C_2} |f(z)|\,|h_{-n-1}(-iz)|\,|dz|. \tag{5.5}$$

Along C_2 we are at a safe distance from the critical points $z=\pm N$ so that formula (3.11) is valid, if we replace κ and K by n and N respectively. It follows that there exists a constant M independent of z and n such that on $E(N, \beta)$ we have

† See E. Hille [1, pp. 435–436] and G. Szegö [7, p. 236].

$$(5.6) \qquad |h_{-n-1}(-iz)| < MA_n |1 - z^2/N^2|^{-1/4} \exp\{-\Im[\xi(z)]\},$$

where

$$(5.7) \qquad A_n = N^{-n-1}e^{n/2} < B(2^n n!)^{-1/2} n^{-1/4},$$

and B is an absolute constant.

On the vertical parts of C_2 it is enough to know that $\Im[\xi(z)]>0$. The length of the vertical parts is $O(n^{-3/4})$ and the factor $|1-z^2/N^2|^{-1/4}$ amounts at most to $O(n^{1/8})$. Their contributions to J_2 are consequently of the form

$$(5.8) \qquad B_2(\beta)[2^n n!]^{-1/2} n^{-7/8} e^{-\beta N}.$$

On the arc PQ of the ellipse, Lemma 2 shows that

$$\int_P^Q |f(z)|\,|h_{-n-1}(z)|\,|dz| < B\,B(\beta)[2^n n!]^{-1/2} n^{-1/4} \cdot \int_P^Q \exp\{-|x|(\beta^2 - y^2)^{1/2} - \Im[\xi(x+iy)]\}\,|1 - z^2/N^2|^{-1/4}\,|dz|$$

$$< B_3(\beta)[2^n n!]^{-1/2} n^{-1/4} e^{-\beta N} \int_P^Q |1 - z^2/N^2|^{-1/4}\,|dz|.$$

The integral being $O(N)$, this expression is of the form

$$(5.9) \qquad B_4(\beta)[2^n n!]^{-1/2} n^{1/4} e^{-\beta N}.$$

Similar estimates hold on C_1. We have then

$$(5.10) \qquad \left| \int_{-N+1}^{N-1} f(t) \boldsymbol{H}_n(t)\,dt \right| < B_5(\beta) n^{1/4} e^{-\beta N},$$

and combining (5.3) and (5.10) we get finally

$$(5.11) \qquad |f_n| < B_6(\beta) n^{1/4} e^{-\beta N}.$$

This being true for every $\beta<\tau$, it is clear that the estimate implies the validity of (5.1). This completes the proof of Theorem 1.

6. **Functions holomorphic or meromorphic in a half-plane.** Suppose that $f(z)$ is holomorphic in the half-plane $y> -\alpha$, $\alpha>0$, and that $f(z)$ admits of a formal Fourier-Hermite series

$$(6.1) \qquad f(z) \sim \sum_{n=0}^{\infty} f_n \boldsymbol{H}_n(z), \qquad f_n = \int_{-\infty}^{\infty} f(t) \boldsymbol{H}_n(t)\,dt.$$

We shall show that if the series has a strip of convergence, then $f(z)$ cannot be of *exponential type* in the half-plane $y \geqq 0$. In other words, the existence of a

strip of convergence implies a lower limit on the rate of growth of the function in its supposed half-plane of holomorphism. The precise statement is as follows.

THEOREM 2. *Let $f(z)$ be holomorphic in the half-plane $y > -\alpha$, $\alpha > 0$. Further let*

$$\limsup_{r\to\infty} (1/r) \log |f(re^{i\theta})| = h(\theta) \leqq M, \qquad 0 \leqq \theta \leqq \pi. \tag{6.2}$$

*The Fourier-Hermite expansion of $f(z)$ can never converge outside of the real axis.**

Proof. For the proof we need the following

LEMMA 3. *If $f(z)$ is holomorphic in the half-plane $y \geqq 0$ and if $\log |f(x)| < -\beta|x|$, where $\beta > 0$, and $\log |f(re^{i\theta})| \leqq Mr$ when $0 \leqq \theta \leqq \kappa$, then $f(z) \equiv 0$.*

This Lemma is implicitly contained in some results of F. Nevanlinna [4, pp. 11–12] and is explicitly stated by F. and R. Nevanlinna [5, p. 38] as a special case of a more general theorem. It is easily proved by forming the Laplace transform of $f(z)$ and rotating the line of integration through an angle of π in the upper half-plane.

This Lemma applies directly to the present situation since the assumption that the Fourier-Hermite series of $f(z)$ has a strip of convergence implies the inequality $\log |f(x)| < -\beta|x|$ for some positive β by virtue of Theorem 1 and this combined with the inequality (6.2) would force $f(z)$ to vanish identically.

We see in particular that *the Fourier-Hermite series of an entire function of order ρ will never converge outside of the real axis unless $\rho > 1$ or $\rho = 1$ and the function is of the maximal type.* Some further results of F. and R. Nevanlinna [5, pp. 40–42] throw light upon the possible width of the strip of convergence of the Fourier-Hermite series of an entire function of order one and particular maximal type. We state the result for functions meromorphic in a half-plane.

THEOREM 3. *Let $f(z)$ be meromorphic in the half-plane $y > -\alpha$, $\alpha > 0$. Let its zeros and poles in the half-plane $y > -\alpha$ be $a_1, a_2, a_3, \cdots$ and $b_1, b_2, b_3, \cdots$ respectively, where in addition $\Im(b_n) \geqq \beta > 0$. Put* $\arg a_n = \alpha_n$, $\arg b_n = \beta_n$, *and let*

$$d = \limsup_{r\to\infty} \frac{1}{r} [b(r) - a(r)],$$

where

$$a(r) = \sum \sin \alpha_n, \qquad b(r) = \sum \sin \beta_n$$

* The same conclusion is valid if the hypotheses refer to a lower half-plane instead. This remark also applies to Theorem 3 below.

and the summations extend over all zeros and poles respectively of absolute value between β and r. Finally let

$$q = \limsup_{r\to\infty} [r \log r]^{-1} \int_0^{\pi} \log \left| f(re^{i\theta}) \right| \sin\theta \, d\theta.$$

If $f(z)$ has a Fourier-Hermite series whose ordinate of convergence equals τ, then $\tau \leqq \min(\alpha, \beta, q+\pi d)$.

In order to connect this theorem with the Nevanlinna theory we have merely to observe that by Theorem 1

$$\limsup_{r\to\infty} \frac{1}{2r} \{\log \left| f(-r) \right| + \log \left| f(r) \right| \} \leqq -\tau$$

and substitution in the inequality of F. and R. Nevanlinna gives $\tau \leqq q+\pi d$. The other inequalities are self-evident.

Theorem 3 shows that the width of the strip of convergence is affected by the frequency of zeros and poles as well as by the rate of growth of the function in parts of the plane distant from the real axis, since all these factors influence the rate of growth of the function near the real axis. That Theorem 3 is the best of its kind is brought out by the following examples.

We first consider $f(z) = \Gamma(a-iz)$, $a>0$. By Stirling's formula and Theorem 1 we have $\tau = \min(a, \pi/2)$. Applying Theorem 3 to $\Gamma(a-iz)$ in the half-plane $y > -a$, we find $\beta = +\infty$, $d=0$, and, again by Stirling's formula, $q=\pi/2$, so that the inequality $\tau \leqq \min(\alpha, \beta, q+\pi d) = \min(a, \pi/2)$ becomes exact. We may of course also apply the theorem to an arbitrary lower half-plane. Then $\beta = a$, $d=1$ and $q=-\pi/2$, so that the same inequality results.

Our second example is $f(z) = \operatorname{sech} az$, $a>0$. Here we can take $\alpha=\beta=\pi/(2a)$ and find $\pi d = a$, $q=0$ and $\tau \leqq \min(a, \pi/(2a))$, which is exact.

References

1. E. Hille, *A class of reciprocal functions*, Annals of Mathematics, (2), vol. 27 (1926), pp. 427–464.

2. ———, *Contributions to the theory of Hermitian series*, Duke Mathematical Journal, vol. 5 (1939), pp. 875–936.

3. R. E. Langer, *On the asymptotic solutions of differential equations, with an application to the Bessel functions of large complex order*, these Transactions, vol. 34 (1932), pp. 447–480.

4. F. Nevanlinna, *Zur Theorie der asymptotischen Potenzreihen*, Annales Academiae Scientiarum Fennicae, (A), vol. 12, no. 3 (1918), vii+81 pp.

5. F. and R. Nevanlinna, *Über die Eigenschaften analytischer Funktionen in der Umgebung einer singulären Stelle oder Linie*, Acta Societatis Scientiarum Fennicae, vol. 50, no. 5 (1922), 46 pp.

6. N. Schwid, *The asymptotic forms of the Hermite and Weber functions*, these Transactions, vol. 37 (1935), pp. 339–362.

7. G. Szegö, *Orthogonal Polynomials*, American Mathematical Society Colloquium Publications, vol. 23, New York, 1939, ix+401 pp.

8. O. Volk, *Über die Entwicklung von Funktionen einer komplexen Veränderlichen nach Funktionen, die einer linearen Differentialgleichung zweiter Ordnung genügen*, Mathematische Annalen, vol. 86 (1922), pp. 296–316.

9. G. N. Watson, *The harmonic functions associated with the parabolic cylinder*, Proceedings of the London Mathematical Society, (2), vol. 8 (1910), pp. 393–421.

10. ———, *The harmonic functions associated with the parabolic cylinder*, Proceedings of the London Mathematical Society, (2), vol. 17 (1918), pp. 116–148.

11. E. T. Whittaker, *On the functions associated with the parabolic cylinder in harmonic analysis*. Proceedings of the London Mathematical Society, vol. 35 (1903), pp. 417–427.

12. E. T. Whittaker and G. N. Watson, *Modern Analysis*, 4th edition, Cambridge, 1927, x+608 pp.

YALE UNIVERSITY,
NEW HAVEN, CONN.

ON 0-REGULAR SURFACE TRANSFORMATIONS*

BY

W. T. PUCKETT, JR.

1. The continuous transformation $T(M)=M'$, where M is a compact metric space, is said to be 0-regular† provided that for each sequence of points $\{x_i'\}$ converging to x' in M', the sets $T^{-1}(x_i')$ converge 0-regularly‡ to $T^{-1}(x')$. This is equivalent to a continuous transformation sending open sets into open sets, while the inverse sets as a collection are uniformly locally connected (that is, for each $\epsilon>0$ a $\delta>0$ exists such that every two points x and y of any inverse set X whose distance apart is less than δ lie in a connected subset of X of diameter less than ϵ). This characterization suggests the projection of a convex euclidean set onto a plane. For example, the orthogonal projection of a solid circle onto a diameter is a 0-regular transformation. It is not 0-regular on the circumference, however, because of the folding about the diameter's end points. That there exist other types of 0-regular transformations is illustrated by the identification of diametrically opposite points of a 2-sphere to obtain a projective plane. A suggestive property of a 0-regular transformation is that the inverse sets must all contain the same number of components.†

In this paper a study is made of 0-regular transformations defined on 2-dimensional pseudo-manifolds. It is shown that if M is a 2-dimensional pseudo-manifold and $T(M)=M'$ is a monotone 0-regular transformation, then either T is topological, or M' is an arc or a simple closed curve. Moreover, it is shown that T must be topological or M' must be degenerate except in the following cases: (i) The sphere, 2-cell, and circular ring may be mapped onto an arc. (ii) The torus, Klein bottle, circular ring, Möbius band, pinched sphere, and 2-cell with two boundary points identified may be mapped onto a simple closed curve. In each of these cases the possible transformations are characterized. For example, it is shown that the only non-topological monotone 0-regular transformation of a sphere onto a nondegenerate image space is equivalent to an orthogonal projection onto a diameter.

* Presented to the Society, April 8, 1939, under the title *The images of 2-dimensional surfaces under 0-regular transformations*; received by the editors August 29, 1939.

† See A. D. Wallace, *On 0-regular transformations*, to appear in the American Journal of Mathematics.

‡ A convergent sequence of closed sets $\{X_n\}$ is said to converge 0-regularly to X provided that for each $\epsilon>0$ there exist positive numbers δ and N such that if $n>N$, any pair of points x, y of X_n with $\rho(x, y)<\delta$ lie together in a continuum in X_n of diameter less than ϵ. See G. T. Whyburn, Fundamenta Mathematicae, vol. 25 (1935), pp. 408–426.

In §6 the above results are stated in terms of possible monotone 0-regular retracting transformations on pseudo-manifolds, while in §7 R. L. Moore's* self-compact equicontinuous collections of curves are used in stating the results. In the concluding section the possible images of pseudo-manifolds under general 0-regular transformations are considered.

2. Throughout this section the following notation will be used: Let M denote a 2-dimensional pseudo-manifold, that is, a 2-dimensional manifold or surface (with or without boundary) among q points of which identifications have been performed so as to produce r local separating points† of M. Let B be the boundary (that is, a finite number of simple closed curves) of M, and denote the finite set of local separating points of M by S. Finally, let $T(M)=M'$ be a *monotone 0-regular* transformation and assume M' *is non-degenerate.*

2.1. *If x is a point of S, then $T^{-1}T(x)=x$.*‡

Proof. Since M is a locally connected continuum and x is a local separating point of M, there exists a connected neighborhood $U(x)$ of x such that $U(x)-x=L_1+L_2+\cdots+L_\lambda$ $(\lambda\geqq 2)$, where the L_i are mutually separated open sets each having x as a limit point. Assume the assertion is false; then there exists an L_k such that $\overline{T^{-1}T(x)\cdot L_k}$ contains x. Let $\{x_i\}$ be a sequence of points in $L_n-T^{-1}T(x)$ (n not k) converging to x. Now $\{T^{-1}T(x_i)\}$ converges to $T^{-1}T(x)$. Hence for each sufficiently large i there exists a point y_i of L_k such that $T(y_i)=T(x_i)$ and $\{y_i\}$ converges to x, since $\overline{T^{-1}T(x)\cdot L_k}$ contains x. But any connected set in $T^{-1}T(x_i)$ containing x_i and y_i must extend outside $U(x)$. Hence $\{T^{-1}T(x_i)\}$ does not converge 0-regularly to $T^{-1}T(x)$ contrary to the hypothesis that T is a 0-regular transformation.

2.11. *If x is a point of S, then $T(x)$ is a local separating point, but not a cut point of M'.*

Proof. Since $T^{-1}T(x)=x$, $T^{-1}T(x)$ locally separates M. Hence by a known theorem§ on monotone transformations $T(x)$ locally separates M'. However, $T(x)$ cannot be a cut point of M' since $M-x$ is connected.

---

* *Foundations of Point Set Theory*, American Mathematical Society Colloquium Publications, vol. 13, New York, 1932, pp. 396–397. The essential definitions are given in §7 for completeness.

† The point x of a continuum M is called a local separating point of M provided there exists a neighborhood $U(x)$ of x in M such that x separates $\overline{U(x)}$ between some pair of points of the component of (Ux) containing x. See G. T. Whyburn, *Local separating points of continua*, Monatshefte für Mathematik und Physik, vol. 36 (1929), pp. 305–314.

‡ The theorem and proof given here are valid if M is any locally connected continuum and x is a local separating point of M.

§ See G. T. Whyburn, *Semi-closed sets and collections*, Duke Mathematical Journal, vol. 2 (1936), p. 686, (3.1).

2.2. *If x' is any point of M', then $T^{-1}(x')$ is either an arc or a simple closed curve.*

Proof. Since T is interior,* $T^{-1}(x')$ can contain no open set. Hence either $T^{-1}(x')$ is a single point (that is, a degenerate arc or simple closed curve) or $T^{-1}(x')$ is a 1-dimensional continuum. Moreover, $T^{-1}(x')$ is locally connected.† Therefore, in order to establish the assertion it must be shown that every point x of $X=T^{-1}(x')$ has an order not greater than 2 in X. Suppose x has an order greater than 2 in X; then there exist nondegenerate arcs α_1, α_2, α_3 in X which are disjoint except for $\alpha_1\cdot\alpha_2\cdot\alpha_3=x$, an end point of each α_i. Since, after 2.1, x cannot be a point of S, there exists a neighborhood $U(x)$ of x in M such that $\overline{U(x)}$ is a 2-cell. It may be assumed each α_i is disjoint‡ with $F(U(x))$ except for the other end point. Now $\overline{U(x)}-\sum\alpha_i$ must contain at least two components L_1, L_2 such that $\overline{L}_i$ contains x. (Observe that one cannot say three components here because x may be a point of B.) Moreover, it may be assumed $\alpha_1\cdot\overline{L}_1=x$, since $\overline{U(x)}$ is a 2-cell. There exists a sequence of points $\{y_i\}$ of $L_1\cdot(M-X)$ converging to x, since X contains no open subset of M. Now if $\{Y_i\}=\{T^{-1}T(y_i)\}$ converges to X, there must exist for all sufficiently large i points z_i of Y_i not contained in L_1, since $\alpha_1\cdot\overline{L}_1=x$. It may be assumed $\{z_i\}$ converges to x. Therefore $\{Y_i\}$ does not converge 0-regularly to X, since any connected subset of Y_i containing y_i and z_i must extend outside $\overline{U(x)}$. Thus the assumption that the order of x is greater than 2 has led to the contradiction that T is not 0-regular.

2.3. *If x' is a point of M' such that $T^{-1}(x')$ is nondegenerate and not contained in B, then x' is a local separating point of M'.*

Proof. Since $X=T^{-1}(x')$ is nondegenerate and not contained in B, it follows from 2.2 and 2.1 that there exists a subarc α of X which is disjoint with $B+S$. Let x be an interior point of α and let α_1, α_2 be arcs such that $\alpha=\alpha_1+\alpha_2$ while $\alpha_1\cdot\alpha_2=x$. There exists a neighborhood $U(x)$ of x which is an open 2-cell and does not contain either of the α_i. Let $x_1xx_2=\beta$ be the subarc of α containing x such that $\beta\cdot F(U(x))=x_1+x_2$; then $U(x)-\beta=L_1+L_2$ is a separation such that $\overline{L}_1\cdot\overline{L}_2=\beta$. Now let $\{U_i(x)\}$ be a sequence of connected neighborhoods contained in $U(x)$ and closing down on x. Then for each i, $U_i(x)-X=L_{i1}+L_{i2}$, where $L_{ij}=L_j\cdot(U_i(x)-X)$, is a separation, since X contains no open subset of M. Suppose x' is not a local separating point of M';

* That is, open sets go into open sets. That a 0-regular transformation is interior follows from the Eilenberg characterization of an interior transformation. See Fundamenta Mathematicae, vol. 24 (1935), p. 174.

† See A. D. Wallace, loc. cit.

‡ For any open set U, $F(U)$ denotes the set-theoretic boundary of U, that is, the set $\bar{U}-U$.

then for each i, $T(U_i(x)-X)=T(U_i(x))-x'=0'$ must be a connected set.* Hence there must exist a point y_i' of $0'$ for each i such that $T^{-1}(y_i')\cdot L_{ij}$ contains a point y_{ij} for $j=1, 2$. Now $\{y_i'\}$ converges to x' and $\rho(y_{i1}, y_{i2})$ converges to zero, since the $U_i(x)$ close down on x. But from the definition of the L_{ij} it follows that any connected set in $T^{-1}(y_i')$ containing y_{i1} and y_{i2} must go outside $U(x)$. Therefore $\{T^{-1}(y_i')\}$ does not converge 0-regularly to X, contrary to the hypothesis that T is a 0-regular transformation.

2.31. *Under the conditions of* 2.3, x' *locally separates* M' *into exactly two components.*

Proof. Suppose x' locally separates M' into more than two components; then there exists a connected neighborhood $V(x')$ such that $V(x')-x'=L_1'+L_2'+L_3'+\cdots$, where the L_j are mutually separated connected open sets with x' a point of $F(L_j')$ for each j. For each j, $T^{-1}(L_j')$ is a connected set, since connectedness is invariant under the inverse of a monotone transformation. Thus, since T is interior, $T^{-1}(x')$ is on the boundary of at least three mutually separated connected open sets in M. This is impossible since $T^{-1}(x')$ is a nondegenerate arc or simple closed curve.

2.32. *If* x' *is an end point of* $T^{-1}(x')$ *for some point* x' *of* M', *then* x *belongs to* B.

2.33. *If* x' *is not a local separating point of* M', *then* $T^{-1}(x')$ *is a single point or is contained in* B.

2.4. *If* x' *is a point of* M' *such that* $B\cdot T^{-1}(x')$ *contains a nondegenerate continuum* K, *then* $T^{-1}(x')$ *is contained in* B *and* x' *is an end point of* M'.

Proof. If $T^{-1}(x')$ is not contained in B, it follows from 2.31 that x' locally separates M' into exactly two components. Thus there exists a connected neighborhood $V(x')$ of x' such that $V(x')-x'=L_1'+L_2'$, where L_1' and L_2' are mutually separated connected open sets with x' a point of $\overline{L_1'}\cdot\overline{L_2'}$. Hence $T^{-1}(L_1')$ and $T^{-1}(L_2')$ are disjoint connected open sets in M whose boundaries have $T^{-1}(x')$ in common, consequently have K in common. This is impossible since $T^{-1}(x')$ is locally connected and K is contained in B.

Let x' be a point of M' such that $T^{-1}(x')$ is nondegenerate and contained in B. Now $T^{-1}(x')$ can contain no point of S. Hence for any point x of $T^{-1}(x')$ which is not an end point there exists a neighborhood $U(x)$ such that $\overline{U(x)}$ is a 2-cell. Moreover, $U(x)$ may be taken so small that x is an interior point of an arc lying in $T^{-1}(x')$ with its end points not in $\overline{U(x)}$.

* See G. T. Whyburn, *Concerning points of a continuous curve defined by certain im kleinen properties*, Mathematische Annalen, vol. 102 (1929), pp. 313–336, Theorem 1.

Thus there exists a neighborhood $V(x')$ such that for each point y_i' of $V(x')-x'$ the set $T^{-1}(y_i')$ contains an arc Y_i, with its end points only in $F(U(x))$, which separates $\overline{U(x)}$ into exactly two components, since T is monotone 0-regular. In case $\{y_i'\}$ is any sequence of points converging to x', the corresponding Y_i may be so selected that if L_i denotes the component of $U(x)-Y_i$ containing x, then x is a point of $L=\prod L_i$ which is contained in $T^{-1}(x')$. Finally, every sufficiently small neighborhood $W(x)$ must have the property that $T^{-1}(x')\cdot W(x)$ is contained in L and for every point y_i' of $T(W(x))-x'$ the product $L_i\cdot W(x)\cdot T^{-1}(y_i')$ is empty. Suppose x' is not an end point of M'. Then it is an interior point of an arc $z_1'x'z_2'$. Now $T^{-1}(z_i'x')$ $(i=1, 2)$ is locally connected.* Hence there exist arcs y_ix_i in $W(x)$ such that $T(y_ix_i)=y_i'x'$ is contained in $z_i'x'$, where it is assumed y_i is the last point of y_ix_i in $T^{-1}(y_i')$ and x_i is the first point in $T^{-1}(x')$. From the choice of $W(x)$ it follows that y_i is contained in Y_i and x_i in L. Since $\overline{U(x)}$ is a 2-cell it may be assumed Y_2 separates Y_1 and L in $U(x)$ and consequently in $W(x)$. Thus y_1x_1 contains a point of Y_2 which contradicts the fact that $T(y_1x_1)$ is contained in $z_1'x'$.

2.5. *If for a point x' of M' the set $T^{-1}(x')$ is not contained in B, then $B\cdot T^{-1}(x')$ can contain only end points of $T^{-1}(x')$.*

Proof. Suppose the assertion is not true. Then there exists a point x of $B\cdot T^{-1}(x')$ which is not an end point of $T^{-1}(x')$. Now x is not a point of S, since $T^{-1}(x')$ must be nondegenerate. Thus there exists a neighborhood $U(x)$ such that $\overline{U(x)}$ is a 2-cell. It may be assumed there exist points of $T^{-1}(x')$ which are not in $\overline{U(x)}$. Let x_1, x_2 be points of $T^{-1}(x')$ different from x such that the arc x_1xx_2 is contained in $U(x)\cdot T^{-1}(x')$, and furthermore, let y, z be points of B different from x such that the arc yxz is contained in $B\cdot U(x)$. Then $(x_1xx_2)\cdot(yxz)=x$; for suppose the product set contained another point x_3. Then x_1xx_2+yxz contains a simple closed curve J, since no subcontinuum of $T^{-1}(x')$ can be contained in B because of 2.4. Thus J separates $\overline{U(x)}$ into exactly two components and there is one component C such that $\overline{C}\cdot[\overline{M-U(x)}]=0$, since J is contained in $U(x)$. Let $\{w_i\}$ be a sequence of points in C converging to a point x_0 of $J\cdot T^{-1}(x')$; then $T^{-1}T(w_i)$ is contained in $\overline{C}=C+J$ for each i. Hence $\{T^{-1}T(w_i)\}$ cannot converge to $T^{-1}T(x_0)$ $=T^{-1}(x')$, since the latter set contains points outside $\overline{U(x)}$. Thus the four arcs x_1x, x_2x, yx, zx are contained in the 2-cell $\overline{U(x)}$ and have by pairs only x in common. Hence there exists a neighborhood $W(x)$ in $U(x)$ which is separated into three components by x_1xx_2 such that x is on the boundary of

* See W. T. Puckett, *Concerning local connectedness . . .*, American Journal of Mathematics, vol. 61 (1939), p. 752, (3.1).

each, and only one can have both x_1x and x_2x on its boundary. Let L_i be the component such that $\overline{L}_i$ contains x_ix but not x_jx, and let $\{w_n\}$ be a sequence of points in L_1 converging to x. Then $\{T^{-1}T(w_n)\}$ cannot converge 0-regularly to $T^{-1}T(x)=T^{-1}(x')$, since they must go outside $W(x)$ to converge to the arc x_2x.

2.6. *If the sequence $\{x_n'\}$ of local separating points of M' converges to a point x' of $M'-T(S)$ and $T^{-1}(x')$ is degenerate, then x' is an end point of M'.*

Proof. There exists a neighborhood $U(x)$ of $x=T^{-1}(x')$ such that $\overline{U(x)}$ is a 2-cell, since x is not a point of S. Moreover, it may be assumed that each $T^{-1}(x_n')$ is contained in $U(x)$, since $\{T^{-1}(x_n')\}$ converges to x. Since x_n' is a local separating point of M', $T^{-1}(x_n')$ locally separates* M and, consequently, separates $\overline{U(x)}$ because it is a 2-cell. But since $T^{-1}(x_n')$ is contained in $U(x)$ it follows that $T^{-1}(x_n')$ separates M; that is, $M-T^{-1}(x_n')=L_n+N_n$ is a separation and it is assumed x is a point of L_n. Now $F(L_n)$ is contained in $T^{-1}(x_n')$, whence $F[T(L_n)]$ contains at most the single point x_n'. Thus $T(L_n)$ is an open set containing x' whose boundary consists of at most a single point. Thus in order to complete the proof it remains to be shown that $T(L_n)$ is a sequence of sets closing down on x'. It may be assumed that for each $n>k$, $T^{-1}(x_n')$ is contained in L_k. Moreover, because the transformation is interior, $F(L_n)=F(N_n)=T^{-1}(x_n')$. Hence x is a point of L_{n+1} which is contained in L_n, and consequently x is a point of $L=\prod L_n$. But $N=\sum N_n$ is open and $F(N)=x$, since for $n<k$, N_n is contained in N_k, $F(N_n)=T^{-1}(x_n')$, and $\{T^{-1}(x_n')\}$ converges to x. Thus if L_n does not close down on x, that is, x is not L, then $M-x=(L-x)+N$ is a separation. But this is impossible, since M is a 2-dimensional pseudo-manifold. Therefore $x=L$, and the proof is complete.

3. We next prove the following theorem.

THEOREM. *If M is a 2-dimensional pseudo-manifold and $T(M)=M'$ is a monotone 0-regular transformation, then T is topological, or M' is either an arc or a simple closed curve.*

Proof. In case M' is degenerate there is nothing to prove. Thus assume M' is nondegenerate and let K be the set of all points of M on which T is one-to-one. Then K is closed, since T is interior. Let $G=M-K$; then by 2.3 and 2.4, $T(G)$ consists of local separating points and end points of M'. Suppose G is not empty, and let x be a point of $F(G)$. If x is not a point of S, then it follows from 2.6 that $T(x)$ is an end point of M'. If x is a point of S, then it follows from 2.11 that $T(x)$ is a local separating point of M'. Hence $T(\overline{G})$ consists of local separating points and end points of M'. Suppose $M-\overline{G}$ is not

* See G. T. Whyburn, *Semi-closed sets and collections*, loc. cit.

empty; then there exist points y of $(M-S)\cdot(M-\overline{G})$ and z of $(M-S)\cdot G$, since $M-\overline{G}$ and G are both open and S is finite. Now $M-S$ is a region in the locally connected continuum M. Hence there exists an arc yz in $M-S$ which must intersect $F(G)$. Let x be the first such point from y to z. Now the arc yx is contained in K, and consequently $T(yx)$ is topological. Therefore no point x_1' of $T(yx)$ can be a local separating point of M', since no point of $M-S$ locally separates M. But $T(x)$ is an end point of M', which is a contradiction. Thus either G is empty or $M-\overline{G}$ is empty. In the first case T is topological, while in the second M' consists of end points and local separating points and consequently is a 1-dimensional continuum. But T is interior and for each point x' of M', $T^{-1}(x')$ is locally connected. Hence it follows from a known theorem* that M' is either an arc or a simple closed curve.

4. It is proposed in this section to show that a monotone 0-regular transformation on a 2-dimensional pseudo-manifold must be topological except in a few specific cases. The notation is that used in §2.

4.1. *In order that T be topological it is sufficient that either* (a) *S contains a point which locally separates M into at least three components, or* (b) *S contains more than one point.*

Proof. Suppose T is not topological; then M' is either an arc or a simple closed curve. Now after 2.11 the image of a point x of S cannot be an end point of M'. Hence $x'=T(x)$ must locally separate M' into exactly two components; that is, there exists a connected neighborhood $V(x')$ in M' such that $V(x')-x'=L'+N'$, where L', N' are open arcs with $x'=\overline{L}'\cdot\overline{N}'$. Suppose x locally separates M into more than two components; then there exists a connected neighborhood $U(x)$ such that $T(U(x))$ is contained in $V(x')$ and $U(x)-x=M_1+M_2+\cdots+M_k$ $(k\geqq 3)$, where, for each integer $i<k$, M_i is a component and x is a point of $\overline{M}_i$. Thus, since $T^{-1}(x')=x$, it may be assumed $L'\cdot T(M_1)\cdot T(M_2)$ contains a sequence of points $\{x_n'\}$ converging to x'. But $\{T^{-1}(x_n')\}$ does not converge 0-regularly to x, since for each n, $T^{-1}(x_n')\cdot M_1$ is not empty and $T^{-1}(x_n')\cdot M_2$ is not empty. Thus for (a) T must be topological. Now under the assumption that T is not topological it follows that if y, z are points of S, then $T(y+z)$ separates M'. Hence $y+z$ separates M, which is impossible. Thus for (b), T is also topological.

The following statement follows immediately from 2.11:

4.11. *In order that M' be an arc it is necessary that S be empty.*

4.2. *If T is not topological, then either $T^{-1}(x')$ is an arc for every point x'*

* See G. T. Whyburn, *Interior transformation on certain curves*, Duke Mathematical Journal, vol. 4 (1938), p. 612.

of M' or every $T^{-1}(x')$ is a simple closed curve. In neither case can more than two $T^{-1}(x')$ be degenerate.

Proof. Let L' be the set of all points x' of M' such that $T^{-1}(x')$ is degenerate. Since T is not topological either x' must be an end point of M' or $x=T^{-1}(x')$ must be a point of S. After 4.11 it follows that S is empty when M' has end points. Thus, since M' is either an arc or simple closed curve and consequently S consists of not more than one point, L' can contain at most two points and $M'-L'$ is connected. Let N_1' be the set of all points x' of M' such that $T^{-1}(x')$ is an arc, and N_2' the set of all points such that $T^{-1}(x')$ is a simple closed curve. From 2.2 it follows that $M'-L'=N_1'+N_2'$. But $\overline{N}_1' \cdot N_2' = N_1' \cdot \overline{N}_2' = 0$ since the 0-regular limit of a sequence of arcs (simple closed curves) is an arc (simple closed curve).* Hence either $N_1'=0$ or $N_2'=0$, since $M'-L'$ is connected.

4.21. *For each point x' of M' let $T^{-1}(x')$ be a simple closed curve, at least one of which is nondegenerate. If J is any simple closed curve of B, then some $T^{-1}(x')=J$.*

Proof. From 2.5 it follows that if $B \cdot T^{-1}(x')$ is not empty then $T^{-1}(x')$ is contained in B. Moreover, it follows from 4.2 that every $T^{-1}(x')$, except possibly two, is a nondegenerate simple closed curve. Thus there exists a nondegenerate simple closed curve $T^{-1}(x')$ contained in B such that $J \cdot T^{-1}(x')$ is not empty. But $J \cdot (\overline{B-J})$ is contained in S (of course, may be empty), while $S \cdot T^{-1}(x')=0$ by 2.1. It follows that $T^{-1}(x')$ is contained in J and is therefore J.

The following assertion also comes out of the above proof:

4.22. *Under the hypotheses of* 4.21, $S \cdot B=0$.

Let $p_m^1(M)$ denote the first Betti number (mod m) of M, for $m \geqq 0$. Also, if N is a closed subset of M, let $p_m^1(N, M)$ denote the first Betti number (mod m) of N relative to M, that is, the number of independent cycles in N relative to homologies (mod m) in M. In another paper† the writer has shown that

(i) *If x', y' are any two points of M', then*

$$p_m^1[T^{-1}(x'), M] = p_m^1[T^{-1}(y'), M]. \tag{1}$$

(ii) *If x' is any point of M', then*

$$p_m^1(M) = p^1(M') + p_m^1[T^{-1}(x'), M], \tag{2}$$

* See G. T. Whyburn, *On sequences and limiting sets*, Fundamenta Mathematicae, vol. 25 (1935), pp. 409–426, particularly (3.1) and (3.2), p. 416.

† *On regular transformations* (offered for publication to Duke Mathematical Journal).

which in the case considered here may be written

$$(2') \qquad 0 \leqq p_m^1(M) - p_m^1(M') = p_n^1[T^{-1}(x'), M],$$

since all the numbers involved are finite.

The two relations above, along with 2.2, give the following assertions:

4.31. *In any case* $0 \leqq p_m^1(M) - p_m^1(M') \leqq 1$ $(m \geqq 0)$.

4.32. *If, for any* $m \geqq 0$, $p_m^1(M) > 2$, *then* T *is topological.*

4.33. *If, for some point* x' *of* M', $T^{-1}(x')$ *is degenerate or an arc, then* $p_m^1(M) = p_m^1(M')$ $(m \geqq 0)$.

4.34. *In order that* M' *be an arc it is necessary that* $p_m^1(M) \leqq 1$ $(m \geqq 0)$.

4.35. *In order that* M' *be a simple closed curve it is necessary that* $1 \leqq p_m^1(M) \leqq 2$ $(m \geqq 0)$.

The assertions 4.1 and 4.11 may be obtained from 4.33, 2.1, and 2.2 as follows: Let M be a pseudo-manifold with $S = y_1 + y_2 + \cdots + y_\lambda$; then it may be assumed M was obtained from a manifold L, which contains no identifications, by identifying μ_i points to obtain y_i. Thus it follows from the Euler-Poincaré formula that

$$(3) \qquad p_m^1(M) = p_m^1(L) + \sum_{i=1}^{\lambda} (\mu_i - 1) \qquad (m \text{ prime}).$$

Therefore, if $\lambda > 1$ or some $\mu_i > 2$ it follows that $p_m^1(M) \geqq 2$. Thus it follows from 2.1 and 4.33 that $p_m^1(M) = p_m^1(M') \geqq 2$, since S is not empty. Hence T must be topological, since M' cannot be an arc of a simple closed curve. By the same reasoning it follows that if S contains a single point then $p_m^1(M') = p_m^1(M) \geqq 1$, and therefore M' cannot be an arc.

A 2-dimensional closed surface M (that is, $S=0$, $B=0$) can possess no 0- or 2-dimensional torsion,* and if M is orientable (that is, a sphere, or torus, and so on) it can possess no 1-dimensional torsion. However, if M is not orientable (that is, projective plane, Klein bottle, and so on) its 1-dimensional torsion group is cyclic of order 2.† Hence it follows from a known theorem‡ that if M is a 2-dimensional closed surface, then

(4a) $p_2^1(M) = p_0^1(M)$ when M is orientable, and

(4b) $p_2^1(M) = p_0^1(M) + 1$ when M is non-orientable.

Therefore it follows† that

(5a) $p_0^1(M) = p_2^1(M) = 0$, if M is a sphere;

* See Alexandroff-Hopf, *Topologie* I, Berlin, 1935, Theorems I and II′, p. 212.

† See *Topologie* I, paragraph 10, pp. 266–269.

‡ See *Topologie* I, Theorem VIII′, p. 227.

(5b) $p_0^1(M)=0$, $p_2^1(M)=1$, if M is a projective plane;

(5c) $p_0^1(M)=1$, $p_2^1(M)=2$, if M is a Klein bottle;

(5d) $p_0^1(M)=p_2^1(M)=2$, if M is a torus; and

(5e) $p_2^1(M)>2$, if M is any other 2-dimensional closed surface.

Now let M be a 2-dimensional surface with boundary (that is, let $S=0$, $B=J_1+J_2+\cdots+J_\beta$, where the J_i are disjoint simple closed curves); then M may be thought of as a closed 2-dimensional surface L with β open 2-cells cut out. Thus if β is not 0, then*

(6a) M possesses no torsion;

(6b) $p_2^1(M)=p_0^1(M)=p_0^1(L)+\beta-1$ when L is orientable; and

(6c) $p_2^1(M)=p_0^1(M)=p_0^1(L)+\beta$ when L is non-orientable.

The relations (3), (5), and (6) are enough to determine the first Betti number (mod 0 or 2) of any 2-dimensional pseudo-manifold with which this paper hereafter is concerned.

4.4. *If M' is a nondegenerate arc, then M must be either a sphere, a 2-cell, or a circular ring.*

Proof. Since M' is an arc, it follows from 4.11 that $S=0$, and from 4.34 that $p_2^1(M)\leqq 1$. Hence, besides those surfaces given in the theorem, M may be either a projective plane or a Möbius band, both of which have $p_2^1(M)=1$. Suppose M to be either one of these and M' to be an arc; then it follows from 4.33 that $T^{-1}(x')$ is a nondegenerate simple closed curve for each point x' of M'. But let y', z' be the end points of M'; then it follows from 2.33 that $T^{-1}(y')+T^{-1}(z')$ is contained in B. This is impossible, since in the first case $B=0$ while in the second B consists of a single simple closed curve.

4.5. *If M' is a nondegenerate simple closed curve, then M is either a torus, a Klein bottle, a circular ring, a Möbius band, a pinched sphere (that is, a sphere with two points identified), or a 2-cell with two boundary points identified.*

Proof. Since M' is a simple closed curve it follows from 4.35 that $1\leqq p_m^1(M)\leqq 2$ for $m=0$ or 2. Thus the torus and Klein bottle are the only closed surfaces which can possibly transform into a simple closed curve. However, if $S=0$ and M has boundary, it is possible, besides the circular ring and Möbius band, that M may be either a 2-cell with two holes, a Möbius band with a hole, a torus with a hole, or a Klein bottle with a hole. In each of these cases $p_2^1(M)=2$ and $\beta\geqq 1$. Suppose M is one of these; then from 4.33 it follows that $T^{-1}(x')$ is a nondegenerate simple closed curve for each point x' of M'. Hence after 4.21 it follows that there exists a $T^{-1}(x')$ contained in B. Thus this x' is an end point of M' because of 2.4. This is impossible under

* See *Topologie* I, paragraph 11, pp. 269–270.

the assumption that M' is a simple closed curve. If S is not empty, then it follows from 2.1 and 4.33 that $p_m^1(M) = p_m^1(M') = 1$, for $m = 0$ or 2. Hence the only possibilities here are the pinched sphere and 2-cell with two points identified. It remains to show that in the case of the 2-cell the two points which are identified must be on the boundary. Let x_1, x_2 be the two points of a 2-cell, at least one of which is not on the boundary, which are identified to give the point $x = S$. Then there exists a neighborhood $U(x)$ such that $U(x) - x$ has two components, one of which, say C, is such that $\overline{C}$ is a 2-cell. Let $\{y_i\}$, contained in C, converge to x. Then for sufficiently large i, $T^{-1}T(y_i)$ is contained in C and is disjoint with $F(C)$. Now each of these $T^{-1}T(y_i)$ must locally separate M and consequently locally separate $\overline{C}$. Thus each must be a nondegenerate simple closed curve. Therefore, $T^{-1}(x')$ is a simple closed curve for each x' of M' after 4.2, and consequently some $T^{-1}(x')$ which is nondegenerate is contained in B. But this x' would be an end point of M' by 2.4, which is impossible since by assumption M' is a simple closed curve.

5. In 4.4 and 4.5 it is shown that only a few of the 2-dimensional pseudomanifolds can possibly be transformed into a nondegenerate arc or simple closed curve by a monotone 0-regular transformation. In this section it is shown that each of these transformations is possible. Moreover, the transformations are completely characterized.

A transformation $T(M) = M'$ is said to be *topologically equivalent** or simply *equivalent* to a transformation $W(N) = N'$ provided one can write $T(M) \equiv hWH(M) = M'$, where $H(M) = N$, $h(N') = M'$ are homeomorphisms.

5.1. *If M is a 2-cell and M' is a nondegenerate arc, then $T(M) = M'$ is equivalent to one of the transformations $W(N) = N'$, where N' is the interval $0 \leqq \xi' \leqq 1$ and either* (a) *N is the square*

$$0 \leqq \xi \leqq 1, \qquad 0 \leqq \eta \leqq 1$$

with $\xi' = \xi$, (b) *N is the triangle*

$$0 \leqq \xi \leqq 1, \qquad 0 \leqq \eta \leqq \xi$$

with $\xi' = \xi$, (c) *N is the triangle*

$$0 \leqq \xi \leqq 1/2, \quad 0 \leqq \eta \leqq \xi; \qquad 1/2 \leqq \xi \leqq 1, \quad 0 \leqq \eta \leqq 1 - \xi$$

with $\xi' = \xi$, or (d) *N is the solid circle*

$$0 \leqq \xi^2 + \eta^2 \leqq 1$$

with $\xi' = (\xi^2 + \eta^2)^{1/2}$.

* See G. T. Whyburn, *Completely alternating transformations*, Fundamenta Mathematicae, vol. 27 (1936), p. 140.

Proof. Suppose, after 4.2, that the inverse of every point of $M'=x_0'x_1'$ is an arc. Then there are three possibilities: (i) neither of the continua $X_i=T^{-1}(x_i')$ $(i=0$ or $1)$ is degenerate, (ii) one, say $X_0=x_0$, is degenerate, and (iii) both $X_0=x_0$ and $X_1=x_1$ are degenerate. It will be shown that the transformations arising from these possibilities are equivalent to (a), (b), and (c) respectively.

Since in (i) X_0 and X_1 are nondegenerate arcs, it follows from 2.33 that they lie in B. Hence $B=\alpha_0+X_0+\alpha_1+X_1$, where each α_i is an arc disjoint with X_0 and X_1 except for one end point in each. Now if x' is an interior point of M', the arc $X=T^{-1}(x')$ must separate X_0 and X_1 in M. Moreover, after 2.5, B can contain only end points of X. Hence $\phi=[X]$, where $X=T^{-1}(x')$ for some x' of M', is an equicontinuous (since it is 0-regular) collection* of arcs satisfying a theorem of R. L. Moore.† Thus there exists a self-compact* collection $G=[g]$ of mutually disjoint arcs such that $\sum g=M$ and for each X of ϕ and g of G the product $X\cdot g$ is a single point. Let $h(N')=M'$ be a topological transformation, and for N of (a) in the theorem let $H_0(M)=N$ be a topological transformation such that $H_0(X_i)=W^{-1}h^{-1}(x_i')$ $(i=0,$ or $1)$. Now $\phi_0=[H_0(X)]$ and $G_0=[H_0(g)]$ are self-compact collections of arcs filling up N. Let $\phi_0'=[X_j^0]$ and $G_0'=[g_j^0]$ be countable subcollections of ϕ_0 and G_0 respectively such that

$$\overline{\sum_{j=1}^{\infty} X_j^0} = N = \overline{\sum_{j=1}^{\infty} g_j^0}.$$

These subcollections may be used in order to set up a sequence $H_j(N)=N$ of topological transformations, each of which is the identity on $\xi=0$, $0\leqq\eta\leqq 1$, whose limit is the homeomorphism $H_\infty(N)=N$ with the property $H_\infty H_0(X)=W^{-1}h^{-1}(x')$ for every point x' of M'. Now define $H\equiv H_\infty H_0$; then $T(M)\equiv hWH(M)$.

In possibility (ii) $X_0=x_0$ is a single point and X_1 is a nondegenerate arc. Just as above X_1 must be contained in B. Moreover, x_0 is contained in B, for suppose it were not. Then there exists a neighborhood $U(x_0)$ disjoint with B and a point x' interior to $x_0'x_1'$ such that $T^{-1}(x')$ is contained in $U(x_0)$ and separates X_1 and x_0 in M. This is impossible, since in the case considered $T^{-1}(x')$ is an arc. Now let $\{y_i'\}$ be a sequence of points converging to x_0'. Then on $M_1=T^{-1}(x_1'y_1')$, and generally on $M_i=T^{-1}(y_{i-1}'y_i')$ the transformation behaves as in (1). Hence just as for (1) there exists for each i a self-compact collection $G^i=[g^i]$ of mutually disjoint arcs such that $\sum g^i=M_i$ and for each $X=T^{-1}(x')$ of M_i, $g^i\cdot X$ is a single point. Now it may be assumed that

* *Foundations of Point Set Theory*, pp. 396–397.

† *Foundations of Point Set Theory*, Theorem 1, p. 397.

for each i, y'_{i+1} precedes y'_i in $x'_0x'_1$. For any point y_1 of $T^{-1}(y'_1)$ let g^1 be the arc of G^1 which has y_1 for an end point, and let g^2 be the arc of G^2 which has y_1 for an end point. Then the other end point of g^2 is a point y_2 of $T^{-1}(y'_2)$. Step by step for each i let g^i be the arc of G^i having y_{i-1} of $T^{-1}(y'_{i-1})$ for one end point and denote the other, which must be in $T^{-1}(y'_i)$, by y_i. Define

$$g = \overline{\sum_{i=1}^{\infty} g^i};$$

then g is an arc. Thus as y_1 ranges over $T^{-1}(y'_1)$ it generates a self-compact collection $G=[g]$ of arcs such that $\sum g=M$ and each g intersects any $X=T^{-1}(x')$ of M in a single point. As for the previous case, this collection along with $\phi=[X]$ may be used in connection with an arbitrary homeomorphism $h(N')=M'$, where N is given by (b), to obtain a topological transformation $H(M)=N$ such that $T(M)\equiv hWH(M)$.

By the argument used above it follows that in possibility (iii) both the points $x_0=X_0$ and $x_1=X_1$ lie in B. Moreover, if x' is an interior point of $M'=x'_0x'_1$, then T behaves on $M_i=T^{-1}(x'_ix')$ as in (ii). Hence if $h(N')=M'$ is an arbitrary homeomorphism, where N and W are given by (c), it follows from (b) that there exist homeomorphisms $H_i(M_i)=W^{-1}h^{-1}(x'_ix')$ such that $T(M_i)\equiv hWH_i(M_i)$. Moreover, the $H_i(M_i)$ may be so defined that $H_0T^{-1}(x')\equiv H_1T^{-1}(x')$. Define $H(x)=H_i(x)$ for x a point of M_i; then $T(M)\equiv hWH(M)$. Thus all three possibilities for $T^{-1}(x')$ an arc are characterized.

If for each x' of M', $T^{-1}(x')$ is a simple closed curve, then it follows from 4.21 and 2.33 that the inverse of one end point of $M'=x'_0x'_1$, say $T^{-1}(x'_1)$, is the whole of B while the inverse of the other end point $T^{-1}(x'_0)=x_0$ is a single point in the interior of M. Since no point separates M, it therefore follows that the collection $\phi=[T^{-1}(x')]$ for all points x' of M' satisfies a theorem of Kerékjártó.* Hence the collection ϕ is homeomorphic with a collection of concentric circles filling a circle. Thus for N and W of (d) there exists a topological transformation $H(M)=N$ such that $WHT^{-1}(x')$ is a point of N' for every point x' of M' and conversely. Thus for each point a' of N' define $h(a')=x'$, where $WHT^{-1}(x')=a'$. Then $h(N')=M'$ is topological and such that $T(M)\equiv hWH(M)$.

5.2. *If M is a sphere and M' is a nondegenerate arc, then $T(M)=M'$ is equivalent to $W(N)=N'$, where N is the sphere $\xi^2+\eta^2+\zeta^2=1$, N is the interval $-1\leqq\xi'\leqq 1$, and W is the transformation $\xi'=\xi$.*

Proof. Since no arc separates the sphere, it follows from 4.2 and 2.33

* See Kerékjártó, *Topologie* I, Berlin, 1933, p. 246. See also H. Whitney, *Regular families of curves*, Annals of Mathematics, (2), vol. 34 (1933), example, p. 260.

that the inverse of every point x' of $M'=x_0'x_1'$ is a nondegenerate simple closed curve except for the end points x_i', which must have degenerate inverses. Let y' be any interior point of M', $M_i'=x_i'y'$, $M_i=T^{-1}(M_i')$, N_i' the interval $[0, (-1)^i]$, and $N_i=W^{-1}(N_i')$ for $i=0$ and 1. Then it follows from 5.1 (d) that there exist homeomorphisms $h_i(N_i')=M_i'$ and $H_i(M_i)=N_i$ such that $T(M_i)\equiv h_iWH_i(M_i)$. It may be assumed $h_0(0)=h_1(0)$ and $H_0T^{-1}(y')\equiv H_1T^{-1}(y')$. Define $H(x)=H_i(x)$ for x a point of M_i and $h(N')=M'$ accordingly. Then $T(M)\equiv hWH(M)$.

5.3. *If M is a circular ring and M' is a nondegenerate arc, then $T(M)=M'$ is equivalent to $W(N)=N'$ where N is the ring $1\leqq\xi^2+\eta^2\leqq 2$, N' is the interval $0\leqq\xi'\leqq 1$, and W is the transformation $\xi'=(\xi^2+\eta^2)^{1/2}-1$.*

Proof. Since $p_2^1(M)$ and $p_2^1(M')$ are not equal, it follows from 4.33 that the inverse of every point x' of M' is a nondegenerate simple closed curve. Let $B=J_0+J_1$, where J_i is a simple closed curve. Then it follows from 4.21 and 2.33 that $J_i=T^{-1}(x_i')$, where $M'=x_0'x_1'$. Let $H_0(M)=N$ be topological and suppose N^* to be the solid circle $0\leqq\xi^2+\eta^2\leqq 1$. Let this be filled with the family of concentric circles $G=[g]$. Then the families G and $\phi=[H_0T^{-1}(x')]$ for all x' of M' satisfy Kerékjártó's condition† that there exists a homeomorphism $H_1(N+N^*)=N+N^*$ such that $[H_1(g)]$ and $[H_1H_0T^{-1}(x')]$ together form a family of concentric circles filling $N+N^*$. Moreover, it may be assumed that for the circle X_0 (that is, for $\xi^2+\eta^2=1$) $H_1(X_0)=X_0$. Define $H\equiv H_1H_0$; then for each x' of M', $WHT^{-1}(x')=a'$, a point of N', and conversely. Now for each point a' of N' define $h(a')=x'$, where $WHT^{-1}(x')=a'$; then $h(N')=M'$ is topological and $T(M)\equiv hWH(M)$.

5.4. *If M is a circular ring and M' is a nondegenerate simple closed curve, then $T(M)=M'$ is equivalent to the transformation $\xi'=\cos\theta$, $\eta=\sin\theta$ on the circular ring N:*

$$\xi = r\cos\theta, \qquad \eta = r\sin\theta \quad (0\leqq\theta\leqq 2\pi,\ 1\leqq r\leqq 2).$$

Proof. Since $S=0$ and M' has no end points, $T^{-1}(x')$ is nondegenerate for each x' of M' and is not contained in $B=J_0+J_1$. Thus, after 4.21, each $T^{-1}(x')$ is a nondegenerate arc, and $B\cdot T^{-1}(x')$ consists of the end points of $T^{-1}(x')$ because of 2.32 and 2.5. Moreover, the end points of $T^{-1}(x')$ must lie one in J_0 and one in J_1, for if both were contained in J_i, $T^{-1}(x')$ would separate M and consequently x' would separate M'. Let y_1' and y_2' be any two points of M'; then $M'=\alpha_1'+\alpha_2'$, where α_i' is an arc and $\alpha_1'\cdot\alpha_2'=y_1'+y_2'$. Define $M_i=T^{-1}(\alpha_i')$. Let $W(N)=N'$ designate the transformation of the theorem and let a_1', a_2' be any two points of N'. Express N' as the sum of

† *Topologie* I, p. 246.

two arcs, $N'=\beta_1'+\beta_2'$ where $\beta_1'\cdot\beta_2'=a_1'+a_2'$. Define $N_i'=W^{-1}(\beta_i')$; then after 5.1 (a) there exist homeomorphisms $h_i(\beta_i')=\alpha_i'$ and $H_i(M_i)=N_i$ such that $T(M_i)\equiv h_iWH_i(M_i)$. Moreover, the homeomorphisms may be so defined that $h_1(a_i')=h_2(a_i')$ and $H_1T^{-1}(y_i')\equiv H_2T^{-1}(y_i')$. Let $H(x)=H_i(x)$ on M_i and $h(a')=h_i(a')$ on β_i; then $T(M)\equiv hWH(M)$.

5.5. *If M is a Möbius band and M' is a nondegenerate simple closed curve, then $T(M)=M'$ is equivalent to the transformation W: $\xi'=\cos\theta$, $\eta'=\sin\theta$, $\zeta'=0$, on the Möbius band N:*

$$\xi=(2+r\cos\theta/2)\cos\theta,\quad \eta=(2+r\cos\theta/2)\sin\theta,\quad \zeta=r\sin\theta/2$$
$$(0\leqq\theta<2\pi,\ -1\leqq r\leqq 1).$$

Proof. Just as in 5.4 it follows that the inverse of each point x' of M' must be a nondegenerate arc with its end points only in B. Again just as in 5.4, 5.1 (a) may be used to define the homeomorphisms h and H such that $T(M)\equiv hWH(M)$, where $W(N)=N'$ is the analytical transformation of the theorem.

5.6. *If M is a 2-cell with two boundary points identified and M' is a nondegenerate simple closed curve, then $T(M)=M'$ is equivalent to the transformation $W(N)=N'$: $\xi'=\cos\theta$, $\eta'=\sin\theta$, where N is defined by*

$$\xi=r\cos\theta,\qquad \eta=r\sin\theta,\qquad 0\leqq\theta\leqq 2\pi,\ 1\leqq r\leqq(1/2)(3-\cos\theta).$$

Proof. Here $S=y_1$ is a single point. Thus it follows from an argument similar to that used in 5.4 that except for $T^{-1}T(y_1)$ the inverse of every point x' of M' is a nondegenerate arc with its end points in B and separated in B by y_1. Express M' as the sum of two nondegenerate arcs, that is, as $M'=\alpha_1'+\alpha_2'$, where $\alpha_1'\cdot\alpha_2'=y_1'+y_2'$. Then 5.1 (b) may be applied here as 5.1 (a) was in 5.4 to give homeomorphisms such that $T(M)\equiv hWH(M)$.

5.7. *If M is a pinched sphere and M' is a nondegenerate simple closed curve, then $T(M)=M'$ is equivalent to the transformation $W(N)=N'$: $\xi'=\cos\theta$, $\eta'=\sin\theta$, $\zeta'=0$, where N is defined by*

$$\xi=(2+\sin^2(\theta/2)\cos\phi)\cos\theta,\quad \eta=(2+\sin^2(\theta/2)\cos\phi)\sin\theta,$$
$$\zeta=\sin^2(\theta/2)\sin\phi\qquad(0\leqq\theta,\phi\leqq 2\pi).$$

Proof. Here again $S=y_1$ is a single point, Just as in 5.2 it follows that the inverse of every point of M', except $T(y_1)$, is a nondegenerate simple closed curve. Express M' as the sum of two nondegenerate arcs one common end point of which is $T(y_1)$. Then 5.1 (d) may be used to define homeomorphisms such that $T(M)\equiv hWH(M)$.

5.8. *If M is a torus and M' is a nondegenerate simple closed curve, then $T(M)=M'$ is equivalent to the transformation $W(N)=N'$: $\xi'=\cos\theta$, $\eta'=\sin\theta$, $\zeta'=0$, where N is defined by*

$$\xi = (2+\cos\phi)\cos\theta, \qquad \eta = (2+\cos\phi)\sin\theta, \qquad \zeta = \sin\phi$$
$$(0 \leqq \theta, \phi \leqq 2\pi).$$

Proof. Since $p_2^1(M) > p_2^1(M')$ it follows from 4.32 and 4.2 that the inverse of every point of M' is a nondegenerate simple closed curve. Let $M'=\alpha_1'+\alpha_2'$, where α_1' and α_2' are nondegenerate arcs with common end points (that is, $\alpha_1'\cdot\alpha_2'=y_1'+y_2'$). Now $M_i=T^{-1}(\alpha_i')$ is a 2-dimensional manifold with $B=T^{-1}(y_1')+T^{-1}(y_2')$. Moreover, $T(M_i)=\alpha_i'$, an arc. Thus M_i must be a circular ring, since this is the only 2-dimensional surface with B consisting of two simple closed curves, which maps into an arc by a monotone 0-regular transformation. Express $N'=\beta_1'+\beta_2'$, where β_1' and β_2' are nondegenerate arcs with end points only in common. Define $N_i=W^{-1}(\beta_i')$. Then 5.3 gives homeomorphisms $h_i(\beta_i')=\alpha_i'$, $H_i(M_i)=N_i$ such that $T(M_i)\equiv h_iWH_i(M_i)$. Moreover, the h_i may be so chosen that $h_1(\beta_1'\cdot\beta_2')\equiv h_2(\beta_1'\cdot\beta_2')$ and the H_i so chosen that $H_1(M_1\cdot M_2)\equiv H_2(M_1\cdot M_2)$. As several times before define $H(x)=H_i(x)$ for x a point of M_i and $h(a')=h_i(a')$ for a' a point of β_i'; then $T(M)\equiv hWH(M)$.

Observation. Let Z_1 and Z_2 be simple closed curves on M which, when oriented, may be considered as generators of the Betti group $B_0^1(M)$. For each pair of positive integers k_1, k_2 there exists a monotone 0-regular transformation of M into a nondegenerate simple closed curve such that for each x' of M' the simple closed curve $T^{-1}(x')$ can be so oriented as to carry a cycle which is homologous to $k_1Z_1+k_2Z_2$. In case $k_i=0$, however, one must choose $k_j=1$.

5.9. While the Klein bottle can be mapped onto a nondegenerate simple closed curve by a monotone 0-regular transformation, $T(M)=M'$, there is not a convenient analytical description as in the previous cases. However, the possible transformations can be characterized after a fashion. In the first place $T^{-1}(x')$, for every point x' of M', must be a nondegenerate simple closed curve, since $p_2^1(M)>p_2^1(M')$. Let Z_1, Z_2 be simple closed curves on M which, when oriented, may be considered as generators of the Betti group $B_0^1(M)$. Since M is a Klein bottle, it may be assumed that $2kZ_2\sim 0$ for all integers k. There exist integers k_1 and k_2 such that

$$T^{-1}(x') \sim k_1Z_1 + k_2Z_2,$$

after $T^{-1}(x')$ is oriented. However k_1 must be zero, for if it were not, then $p_0^1[T^{-1}(x'), M]=1$. Consequently

$$p_0^1(M') = p_0^1(M) - p_0^1[T^{-1}(x'), M] = 0,$$

contrary to the fact that M' is a nondegenerate simple closed curve. Thus for every x' of M', $T^{-1}(x') \sim k_2 Z_2$, when oriented. Moreover, k_2 must be odd, since $p_2^1[T^{-1}(x'), M]$ cannot be zero.

In order to demonstrate such a mapping suppose M to arise from the oriented square $ABCD$ by identifying the oriented sides, AB with DC and BC with DA.* Let $T(M) = M'$ be such that the collection $[T^{-1}(x')]$ in $ABCD$ is a collection of straight lines parallel to BC.

6. The continuous transformation $T(M) = N$, a subset of M, is said to be *retracting*† provided that for each point x of N, $T(x) = x$. The following statements are immediate consequences of the results in the preceding sections:

6.1. *In order that there exist a monotone* 0*-regular retracting transformation of the* 2*-dimensional pseudo-manifold* M *onto a nondegenerate arc, it is necessary and sufficient that* M *be a* 2*-cell, a circular ring, or a sphere.*

6.2. *In order that there exist a monotone* 0*-regular retracting transformation of the* 2*-dimensional pseudo-manifold* M *onto a nondegenerate simple closed curve, it is necessary and sufficient that* M *be a circular ring, a Möbius band, a torus, a Klein bottle, a* 2*-cell with two boundary points identified, or a pinched sphere.*

6.3. *There exist no monotone* 0*-regular retracting transformations of* 2*-dimensional pseudo-manifolds onto nondegenerate sets except those given by* 6.1 *and* 6.2.

7. A collection G of continua is said to be *equicontinuous*‡ with respect to a given set M if for every collection H of open sets covering M there exists a finite collection H' of open sets covering M such that if x_1 and x_2 are two points of M lying in some one set of H' and belonging to a continuum X of G, then there exists an arc x_1x_2 lying both in X and in some set of the collection H. The collection G is said to be *self-compact*‡ if every infinite sequence of continua of the collection G contains an infinite subsequence which converges to some set of the collection G.

Let M be any compact space and $T(M) = M'$ be a monotone 0-regular transformation. Then, obviously, the collection $G = [T^{-1}(x')]$ for all points x' of M' is an equicontinuous self-compact collection of mutually disjoint continua filling M. Moreover, it is easily seen that an equicontinuous self-compact collection $G = [X]$ of mutually disjoint continua filling M gives rise

* See Alexandroff-Hopf, p. 207.

† See Borsuk, Fundamenta Mathematicae, vol. 18 (1932), p. 204.

‡ *Foundations of Point Set Theory*, pp. 396–397.

to a monotone 0-regular transformation. Thus the following assertions are immediate consequences of the results of §§4 and 5:

7.1. *Let M be a 2-dimensional pseudo-manifold and $G=[X]$ be an equicontinuous self-compact collection of mutually disjoint continua filling M. If G contains more than one element, then each X is an arc or each X is a simple closed curve.*

7.2. *In order that a 2-dimensional pseudo-manifold M may be decomposed into an equicontinuous self-compact collection of mutually disjoint arcs, at least one of which is nondegenerate, it is necessary and sufficient that M be a 2-cell, a 2-cell with two boundary points identified, a circular ring, or a Möbius band.*

7.3. *In order that a 2-dimensional pseudo-manifold M may be decomposed into an equicontinuous self-compact collection of mutually disjoint simple closed curves, at least one of which is nondegenerate, it is necessary and sufficient that M be a 2-cell, a circular ring, a sphere, a pinched sphere, a torus, or a Klein bottle.*

8. A. D. Wallace* has shown that if $T(M)=T_2T_1(M)$ is the usual monotone-light factoring of any 0-regular transformation, then T_1 is a monotone 0-regular transformation and T_2 is a local homeomorph sm.† Moreover, he has shown that when the image is nondegenerate any 0-regular transformation on an arc or on a simple closed curve is a homeomorphism or a local homeomorphism respectively. Thus the following assertions are consequences of the results of §§3, 4, and 5:

8.1. *If M is a 2-dimensional pseudo-manifold and $T(M)=M'$ is a 0-regular transformation, then M' is a 2-dimensional pseudo-manifold, an arc, or a simple closed curve.*

8.2. *Let M be a 2-dimensional pseudo-manifold and $T(M)=M'$ a 0-regular transformation. If M' is a nondegenerate arc, then T is monotone and, consequently, is equivalent to one of the transformations in* 5.1, 5.2, *or* 5.3.

Since a local homeomorphism on a simple closed curve is equivalent to the transformation $W'\colon \xi'=\cos k\theta, \eta'=\sin k\theta$ (k an integer) on the circle $\xi=\cos\theta$, $\eta=\sin\theta$ $(0\leqq\theta\leqq 2\pi)$,‡ the following assertion is immediate:

---

* *On 0-regular transformations*, loc. cit.

† The transformation $T(M)=M'$ is said to be a local homeomorphism if for each point x of M there exists a neighborhood $U(x)$ on which T is topological. See S. Eilenberg, Fundamenta Mathematicae, vol. 24 (1935), p. 35.

‡ See G. T. Whyburn, *Interior transformations on compact sets*, Duke Mathematical Journal, vol. 3 (1937), p. 374, (3.2).

8.3. *Let M be a 2-dimensional pseudo-manifold and $T(M) = M'$ a 0-regular transformation. If M' is a nondegenerate simple closed curve, then T is equivalent to the transformation $W'W$, where W' is given above and W is one of the transformations in* 5.4, 5.5, 5.6, 5.7, 5.8, *or* 5.9.

The following assertion results from 6.3, 8.2, and 8.3:

8.4. *If M is a 2-dimensional pseudo-manifold and $T(M) = N$ is a 0-regular retracting transformation, then T must be monotone.*

UNIVERSITY OF VIRGINIA,
CHARLOTTESVILLE, VA.

INTEGRATION IN A CONVEX LINEAR TOPOLOGICAL SPACE*

BY

R. S. PHILLIPS

This paper is concerned with a theory of integration for functions with values in a convex linear topological space. We consider an integral which is essentially an extension to this general space of the integral studied by Garrett Birkhoff [1] in a Banach space. By imposing different convex neighborhood topologies on a Banach space, we obtain as instances of our integral those defined by Birkhoff [1], Dunford [2], Gelfand [3], and Pettis [4].

Let $f(s)$ be a function on an abstract set S to the real numbers, and let $\alpha(\sigma)$ be a nonnegative measure function on an additive family of subsets $\mathfrak{S}$ of S. A necessary and sufficient condition for the Lebesgue integral to exist is that for each $e>0$ there exist a partition Δ_e of S into a denumerable set of sets (σ_i) such that for any two orderings of these sets, (σ_i^1) and (σ_i^2),

$$\left| \limsup_n \sum^n \alpha(\sigma_i^1) \cdot \sup_{s_i \in \sigma_i^1} f(s_i) - \liminf_n \sum^n \alpha(\sigma_i^2) \cdot \inf_{s_i \in \sigma_i^2} f(s_i) \right| < e.$$

We have introduced upper and lower limits in order to stress the fact that unconditional convergence is an unnecessary concept in the definition of the Lebesgue integral. In fact by avoiding unconditional convergence we have been able to extend this integral to a linear convex topological space. This further permits a simplification in the concepts involved.

In the first section we consider a few relations between linear continuous operations and convex neighborhoods which have been found useful in the study of the integral. In §2 the integral is defined and some essential properties of unconditional summability are established. This is followed in §3 by a discussion of the basic properties of the integral. The integral of an integrable function x is an absolutely continuous and completely additive set function depending linearly on x. It is found in §4 that the integrability conditions can be relaxed when the space satisfies a certain type of completeness condition. A demonstration that the linear continuous transform of an integrable function is itself integrable occupies §5. Section 6 treats convergent sequences of integrable functions, while §7 deals with relations to other integrals. A theorem on differentiation is proved in §8. In §9 an application is made of this theory to an existence theorem in differential equations. Finally in §10

* Presented to the Society, April 15, 1939; received by the editors June 19, 1939.

some examples are given to demonstrate properties of the integral. Example 10.2 shows that the Birkhoff and Pettis integrals are different. As a consequence our development leads to the definition of a class of integrals for the theory of integration on a Banach space.

1. **On convex linear topological spaces.** We will restrict ourselves in this paper to the type of convex linear topological space X defined by Kolmogoroff [5, p. 29] to satisfy the following axioms:

1. X is a linear space relative to real numbers [6, p. 26].

2. There exists an operation of closure, $\overline{B}$, for any subset B of X with the following properties:

2a. If B consists of a single element of X, then $\overline{B}=B$.

2b. $\overline{B}=\overline{\overline{B}}$.

2c. $\overline{B+C}=\overline{B}+\overline{C}$ ($+$ stands for set addition).

3. Addition of elements and multiplication by numbers are continuous.

4. There exists an equivalent neighborhood system of convex* neighborhoods $\mathcal{U}$. Kolmogoroff has shown that Axioms 1, 2, 3 are sufficient to make X a regular Hausdorff space with a uniform topology.† We will therefore use only neighborhoods of the null vector.

In a convex space *pseudo-norms* have been defined by von Neumann [7, pp. 18–19] as follows: Let $\|x\|_U^+$ equal the greatest lower bound of $a>0$ such that $x \in aU$, and let $\|x\|_U=\max[\|x\|_U^+, \|-x\|_U^+]$. He has established the following properties by arguments that are applicable to our space:

$$\|x+y\|_U \leqq \|x\|_U + \|y\|_U, \quad \|ax\|_U = |a| \cdot \|x\|_U, \quad \|\theta\|_U = 0,$$

$\|x\|_U$ is a continuous function of x, and the sets $\|x\|_U<\delta$, $U \in \mathcal{U}$, $\delta>0$ form a complete system of neighborhoods for θ. The norm has the additional property: $\|x\|=0$ if and only if $x=\theta$. This is not a property of the pseudo-norm. In the remainder of this paper we will use the system of neighborhoods defined by the pseudo-norms and designate by U the set of all x such that $\|x\|_U<1$. Hence $U_1 \supset U_2$ is equivalent to the statement that $\|x\|_{U_2}<1$ implies $\|x\|_{U_1}<1$. Since the pseudo-norm is homogeneous and continuous, this is in turn equivalent to $\|x\|_{U_1} \leqq \|x\|_{U_2}$.

Let us now set up a one-to-one correspondence between the set of all linear continuous operations‡ on X and an abstract class T of the same cardinal power; say, $\bar{x}_t$ corresponds to $t \in T$. We proceed to prove that the neighborhoods of X can be defined by linear continuous operations on X.

---

* A set B is convex if, whenever $x, y \in B$, $a, b \geqq 0$, and $a+b=1$, then $ax+by \in B$.

† X is said to have a *uniform topology* if, when $\mathcal{U}(x)$ and $\mathcal{U}(\theta)$ are the neighborhood systems for x and θ respectively, $x+\mathcal{U}(\theta)$ is equivalent to $\mathcal{U}(x)$.

‡ The term "operation" is used for a numerical function.

THEOREM 1.1. *For each $U \in \mathcal{U}$ there exists a subset T_U of T such that* $[x \mid x \in X, |\bar{x}_t(x)| < 1 \text{ for every } t \in T_U] \equiv [x \mid x \in X, \|x\|_U < 1]$.*

Wehausen [8, p. 162, Theorem 8] has demonstrated that given any $x_0 \in X$ and $U \in \mathcal{U}$, there exists a linear continuous operation $\bar{x}_0$ on X with the property that $|\bar{x}_0(x)| \leqq \|x\|_U$ and $\bar{x}_0(x_0) = \|x_0\|_U$. To each $x \in X$ corresponds in this fashion $\bar{x}_{t_x}$. To the set of $x \in X$ for which $\|x\|_U = 1$ will then correspond a subset $T_U \equiv [t_x]$ of T. For $t \in T_U$, $|\bar{x}_t(x)| \leqq \|x\|_U$, and hence $[x \mid x \in X, |\bar{x}_t(x)| < 1$ for every $t \in T_U] \supset [x \mid x \in X, \|x\|_U < 1]$. Further if $|\bar{x}_t(x)| < 1$ for every $t \in T_U$, then either $\|x\|_U = 0$, or $\|x\|_U \neq 0$, in which case there exist $t_{ax} \in T_U$ and $a > 1$ such that $|\bar{x}_{t_{ax}}(ax)| = 1 = \|ax\|_U$ and $\|x\|_U = 1/|a| < 1$. The inclusion therefore holds both ways.

We will use interchangeably the three symbols U, $[x \mid x \in X, \|x\|_U < 1]$, and $[x \mid x \in X, |\bar{x}_t(x)| < 1$ for every $t \in T_U]$ since they represent the same subset of X. The last expression can be considered as an analysis of the neighborhood U just as the neighborhoods U furnish a decomposition of the neighborhood system $\mathcal{U}$.

It is instructive to investigate the relation between the set of all linear continuous operations on X and that subset used to define the neighborhood system. This can be conveniently described by means of the algebraic notion of a factor group. Essentially the factor group permits us to study a linear convex topological space by means of a set of linear normed vector spaces.

If Θ is a closed linear manifold in X, then, being a subgroup of X, Θ defines a factor group Y whose elements y correspond biuniquely to the cosets. Y is again a linear space. Now for a given $U \in \mathcal{U}$ we will define $\Theta_U \equiv [x \mid x \in X, \|x\|_U = 0]$ or what is equivalent $\Theta_U \equiv [x \mid x \in X, \bar{x}_t(x) = 0$ for every $t \in T_U]$. Then as Θ_U is linear and closed, it defines the factor group Y_U. We define a norm on Y_U to be $\|y\| = \|x\|_U$ where x is any element of the coset y. This is clearly independent of the particular $x \in$ coset y used and satisfies all of the properties of a norm. The transformation $y = T(x)$, where $x \in$ coset y, is a linear continuous transformation on X to Y_U with its norm topology. A linear continuous operation $\bar{y} \in \overline{Y}_U$, the space adjoint to Y_U, defines a linear continuous operation $\bar{x} = \overline{T}(\bar{y})$, namely $\bar{x}(x) = \bar{y}(y)$ where $x \in$ coset y. We will use this correspondence in the statement of the following theorem:

THEOREM 1.2. *The set of all linear continuous operations on X is the linear extension of the set of operations which correspond to the adjoint spaces $\overline{Y}_U$ for all $U \in \mathcal{U}$.*

If $\bar{y} \in \overline{Y}_U$, then $|\bar{x}(x)| = |\bar{y}(y)| \leqq \|\bar{y}\| \cdot \|y\| = \|\bar{y}\| \cdot \|x\|_U$. In other words

* We will designate by $[y \mid P]$ the class of elements y with the properties P.

$|\bar{x}(x)| < e$ if $x \in eU/\|\bar{y}\|$. On the other hand if $\bar{x}$ is continuous on X, then by a theorem due to Wehausen [8, pp. 162–163, Theorem 9], there exists a $U \in \mathcal{U}$ and a positive constant M for which $|\bar{x}(x)| \leqq M \cdot \|x\|_U$. The correspondence $\bar{y}(y) = \bar{x}(x)$ where $x \in$ coset y makes $\bar{x}$ correspond to an element of $\overline{Y}_U$.

THEOREM 1.3. *If the set* $[\bar{x}_t | t \in T_U]$ *has* n *and only* n *linearly independent elements, then* Y_U *has* n *and only* n *linearly independent elements.*

Suppose first that there exist $n+1$ linearly independent vectors $y_1, \cdots, y_{n+1}$ in Y_U, and that $\bar{x}_1, \cdots, \bar{x}_n \in [\bar{x}_t | t \in T_U]$ are linearly independent. If $\bar{y}_t$ corresponds to $\bar{x}_t$, then the system of n homogeneous linear equations $\sum_{i=1}^{n+1} a_i \bar{y}_t(y_i) = 0$ $(t = 1, \cdots, n)$ has a non-trivial solution $\{a_i\}$. Let x_i be an element of coset y_i. Since $\sum_{i=1}^{n+1} a_i \bar{x}_t(x_i) = \sum_{i=1}^{n+1} a_i \bar{y}_t(y_i) = 0$ $(t = 1, \cdots, n)$ and since any other $\bar{x}_t$ for $t \in T_U$ is a linear combination of $\bar{x}_1, \cdots, \bar{x}_n$, it follows that $\bar{x}_t(\sum_{i=1}^{n+1} a_i x_i) = 0$ for all $t \in T_U$. Hence $\sum_{i=1}^{n+1} a_i x_i \in \Theta_U$ which is contrary to our supposition. On the other hand, if there do not exist n linearly independent elements in Y, then some linear combination of every set of n elements of X belongs to Θ_U. Hence for every set $x_1, \cdots, x_n$, the n rowed determinant $|\bar{x}_t(x_j)| = 0$ $(t, j = 1, \cdots, n)$. Let r be the largest rank attained by $|\bar{x}_t(x_j)|$ for any choice of $\{x_j\}$—say $|\bar{x}_t(x_j)| \neq 0$ $(t, j = 1, \cdots, r)$. Then $|\bar{x}_t(x_j)| = 0$ $(t, j = 1, \cdots, r+1)$ for any $x_{r+1} \in X$ and hence there exist $\{c_t\}$ such that $\sum_{t=1}^{r+1} c_t \bar{x}_t(x) = 0$ for all $x \in X$. But this means that $\bar{x}_1, \cdots, \bar{x}_{r+1}$ are linearly dependent.

COROLLARY 1.4. *If each of the sets* $[\bar{x}_t | t \in T_U]$ *has only a finite number of linearly independent elements, then the set of continuous linear operations on* X *is just the linear extension of* $[\bar{x}_t | t \in T_U, U \in \mathcal{U}]$.

Since an n-dimensional normed vector space has an n-dimensional adjoint space, this is an immediate consequence of Theorems 1.2 and 1.3.

If a weak topology is defined on a linear vector space by means of all finite subsets of a certain class X^* of linear operations on the space, then, by the corollary, the set of all linear continuous operations on this space is exactly the linear extension of X^*.

2. **The integral and unconditional summability.** Let S be an abstract set possessing a sigma-field $\mathfrak{S}$ of "measurable" subsets σ. We will study the integration of multiple-valued set functions $x(\sigma)$ having the property that the set $x(\sigma_1)$ contains the set $x(\sigma_2)$ if $\sigma_1 \supset \sigma_2$. A point function $x(s)$ on S to X defines a set function of this type $x(\sigma)$ where $x(\sigma)$ denotes the set of all elements $x(s)$ for $s \in \sigma$. The function $\alpha(\sigma)$ will be a single-valued, nonnegative, bounded, completely additive measure function on $\mathfrak{S}$. We will designate by Δ a subdivision of S into a finite or denumerable set of sets (σ_i) such that

$\sigma_i \in \mathfrak{S}$, $\sigma_i \cdot \sigma_j = 0$ if $i \neq j$, and $\sum \sigma_i = S$. The expression $\Delta_1 \geqq \Delta_2$ will mean that every set σ^1 of Δ_1 is a subset of some σ^2 of Δ_2, and $\Delta_1 \cdot \Delta_2$ will be the subdivision of sets $(\sigma_i^1 \cdot \sigma_j^2)$, σ^i of Δ_i. The convex extension of the subset B of X which is the set of all $\sum_1^n a_i x_i$ where $a_i \geqq 0$, $\sum_1^n a_i = 1$, $x_i \in B$, will be written as cvx (B). π will be a finite set of positive integers. $\pi_1 \geqq \pi_2$, $\pi_1 + \pi_2$, $\pi_1 \cdot \pi_2$, and $\pi_1 - \pi_2$ will denote respectively inclusion, the integers in π_1 or π_2, the integers in π_1 and π_2, the integers in π_1 not in π_2. $\sum_\pi B_i$ will be the set of all $\sum_\pi x_i$ summed over π where x_i is an element of the subset B_i of X. Let $\{B_i\}$ be a finite or denumerable sequence of subsets of X. We will designate by $[\sum_\pi B_i]$ the set of all sums $\sum_\pi B_i$. In this notation $[\sum_\pi B_i]$ will be called unconditionally summable with respect to U if there exists π_U such that if $\pi_i \geqq \pi_U$ $(i = 1, 2)$, then $\|\sum_{\pi_1} B_i - \sum_{\pi_2} B_i\|_U < 1$;† and unconditionally summable to the value $x \in X$ with respect to U if there exists π_U such that if $\pi \geqq \pi_U$, then $\|\sum_\pi B_i - x\|_U < 1$. If the B_i are real numbers, then unconditional summability of $[\sum_\pi B_i]$ is equivalent to the absolute convergence of $\sum_{i=1}^\infty B_i$ [9, p. 63].

We are now in a position to define the $\mathcal{U}$-integral. Let $J_\Delta(x, \sigma) = [\sum_\pi x(\sigma \cdot \sigma_i)\alpha(\sigma \cdot \sigma_i)]$ where $\Delta = (\sigma_i)$.

DEFINITION 2.1. *$x(\sigma)$ will be said to be $\mathcal{U}$-integrable if for every $\sigma \in \mathfrak{S}$ there exists a $J(x, \sigma) \in X$ such that given $U \in \mathcal{U}$ there exists $\Delta_{U\sigma}$ for which $J_{\Delta_{U\sigma}}(x, \sigma)$ is unconditionally summable to the value $J(x, \sigma)$ with respect to U. We define $J(x, \sigma)$ to be the value of the integral on the set σ; that is, $J(x, \sigma) = \int_\sigma x d\alpha$.*

The uniqueness of $J(x, \sigma)$ will be a consequence of Theorem 2.3 to follow.

The essential difference exhibited by this definition is that it does not require the sums $\sum_1^\infty x_i \alpha(\sigma_i)$ where $x_i \in x(\sigma_i)$ to converge. Instead it requires the sums of a subdivision corresponding to a given neighborhood U to be contained in $J(x, \sigma) + U$ whenever the set of integers over which the sum is taken includes a certain finite set. We are thus freed from convergence problems. On the other hand to have demanded the unconditional convergence of the sums $\sum x(\sigma_i)\alpha(\sigma_i)$ would have placed unnecessary restrictions either on $x(\sigma)$ or on X. Finally it has been necessary to postulate the existence of $J(x, \sigma)$ because a linear topological space need not be complete.

THEOREM 2.2. *If, for a given Δ, $J_\Delta(x, S)$ is unconditionally summable with respect to U, then $J_\Delta(x, \sigma)$ is unconditionally summable with respect to $3U$ uniformly in σ (that is, π_{3U} is the same for all σ).*

By hypothesis there exists π_U such that if $\pi_i \geqq \pi_U$ $(i = 1, 2)$, then $[\sum_{\pi_1} x(\sigma_i)\alpha(\sigma_i) - \sum_{\pi_2} x(\sigma_i)\alpha(\sigma_i)] \subset U$. Since U is convex, we may write the relationship

† If B is a subset of X, then $\|B\|_U < 1$ means that $B \subset U$.

$$U \supset \frac{\alpha(\sigma \cdot \sigma_{i_1})}{\alpha(\sigma_{i_1})}\left[\sum_{\pi_U}[x(\sigma_i)\alpha(\sigma_i) - x(\sigma_i)\alpha(\sigma_i)]\right] + \left[1 - \frac{\alpha(\sigma \cdot \sigma_{i_1})}{\alpha(\sigma_{i_1})}\right]\left[\sum_{\pi_U}[x(\sigma_i)\alpha(\sigma_i) - x(\sigma_i)\alpha(\sigma_i)]\right]$$

where $i_1 \in \pi_U$. Let us take any element of $[x(\sigma_i)\alpha(\sigma_i) - x(\sigma_i)\alpha(\sigma_i)]$ in the first term on the right and take the same element of $[x(\sigma_i)\alpha(\sigma_i) - x(\sigma_i)\alpha(\sigma_i)]$ in the second except in the case of i_1 where we always take $\theta \in [x(\sigma_{i_1})\alpha(\sigma_{i_1}) - x(\sigma_{i_1})\alpha(\sigma_{i_1})]$ in the second term. This will give

$$U \supset \sum_{\pi_U - i_1}[x(\sigma_i)\alpha(\sigma_i) - x(\sigma_i)\alpha(\sigma_i)] + [x(\sigma_{i_1})\alpha(\sigma \cdot \sigma_{i_1}) - x(\sigma_{i_1})\alpha(\sigma \cdot \sigma_{i_1})].$$

In a like manner we obtain

$$U \supset \sum_{\pi_U - i_1 - i_2}[x(\sigma_i)\alpha(\sigma_i) - x(\sigma_i)\alpha(\sigma_i)] + \sum_{i_1 + i_2}[x(\sigma_i)\alpha(\sigma \cdot \sigma_i) - x(\sigma_i)\alpha(\sigma \cdot \sigma_i)]$$

where $i_2 \in \pi_U$. Treating successively the integers of π_U, we will finally arrive at

$$(1) \quad U \supset \sum_{\pi_U}[x(\sigma_i)\alpha(\sigma \cdot \sigma_i) - x(\sigma_i)\alpha(\sigma \cdot \sigma_i)] \supset \sum_{\pi_U}[x(\sigma \cdot \sigma_i)\alpha(\sigma \cdot \sigma_i) - x(\sigma \cdot \sigma_i)\alpha(\sigma \cdot \sigma_i)].$$

If $\pi_0 \cdot \pi_U = 0$, then

$$(2) \quad \sum_{\pi_0} x(\sigma \cdot \sigma_i)\alpha(\sigma \cdot \sigma_i) \subset \sum_{\pi_0} x(\sigma_i)\alpha(\sigma \cdot \sigma_i) \subset \sum_{\pi_0} \operatorname{cvx}[x(\sigma_i)\alpha(\sigma_i) + \theta] \subset \operatorname{cvx}\left[\sum_{\pi} x(\sigma_i)\alpha(\sigma_i) \,\middle|\, \text{finite set of } \pi\text{'s where } \pi \cdot \pi_U = 0\right] \subset U.$$

(1) and (2) together give $(\sum_{\pi_1} x(\sigma \cdot \sigma_i)\alpha(\sigma \cdot \sigma_i) - \sum_{\pi_2} x(\sigma \cdot \sigma_i)\alpha(\sigma \cdot \sigma_i)) \subset 3U$ for all $\sigma \in S$ where $\pi_i \geqq \pi_U$ $(i = 1, 2)$.

THEOREM 2.3. *If $J_{\Delta_0}(x, \sigma)$ is unconditionally summable to $J(x, \sigma)$ with respect to U, then for any $\Delta \geqq \Delta_0$, $J_\Delta(x, \sigma)$ is unconditionally summable to $J(x, \sigma)$ with respect to $4U$.*

Without loss of generality we may consider only the case $\sigma = S$. By hypothesis there exists π_0 such that if $\pi \geqq \pi_0$, then $(\sum_\pi x(\sigma_i)\alpha(\sigma_i) - J(x, S)) \subset U$ where $\Delta_0 = (\sigma_i)$. Let $(\sigma_j{}^i)$ be a subdivision of σ_i. Then $\Delta = (\sigma_j{}^i)$ $(i, j = 1, 2, \cdots)$ is a subdivision of S. We first propose to prove a lemma.

LEMMA. *Let $x(\sigma_i)\alpha(\sigma_i) + Y_i \subset b_i U$ where Y_i is a subset of X and $b_i > 0$. Then there exists an n_i such that if $n \geqq n_i$ then*

$$\left\| \left[\sum_{j=n+1}^{\infty} \alpha(\sigma_j{}^i)\right] x(\sigma_i) \right\|_U < \frac{1}{2^i}$$

and

$$\sum_{j=1}^{n} x(\sigma_j{}^i)\alpha(\sigma_j{}^i) + Y_i \subset b_i U + \frac{1}{2^i} U.$$

Since U is convex,

$$b_iU \supset \operatorname{cvx}\,[x(\sigma_i)\alpha(\sigma_i) + Y_i] \supset \frac{1}{c_n^i}\sum_{j=1}^{n} x(\sigma_j^i)\alpha(\sigma_j^i) + Y_i$$

where $c_n^i = \sum_{j=1}\alpha(\sigma_j^i)/\alpha(\sigma_i) \leqq 1$. Hence

$$\sum_{j=1}^{n} x(\sigma_j^i)\alpha(\sigma_j^i) + c_n^i Y_i \subset c_n^i b_i U_i \subset b_i U.$$

As $\|x(\sigma_i)\alpha(\sigma_i)+Y_i\|_U < b_i$, there exists an M_i for which $\|x(\sigma_i)\alpha(\sigma_i)\|_U + \|Y_i\|_U < M_i$. Now $c_n^i \to 1$ as $n \to \infty$ so that there will be an n_i such that if $n \geqq n_i$ then $\|(1-c_n^i)Y_i\|_U < 1/2^i$ and $\|(1-c_n^i)x(\sigma_i)\alpha(\sigma_i)\|_U < 1/2^i$. Therefore $\sum_{j=1}^{n} x(\sigma_j^i)\alpha(\sigma_j^i) + Y_i < b_iU + U/2^i$ for $n \geqq n_i$.

We will now apply this lemma successively to the elements of π_0. If i_n is the nth integer of π_0 so chosen, then

$$Y_{i_n} = \sum_{\pi_0 - \pi_n - i_n} x(\sigma_i)\alpha(\sigma_i) + \sum_{\pi_n}\left[\sum_{j=1}^{n_i} x(\sigma_j^i)\alpha(\sigma_j^i)\right] - J(x, S),$$

and $b_{i_n} = \sum_{0+\pi_n} 1/2^i$ where $\pi_n = (i_1, i_2, \cdots, i_{n-1})$. We finally obtain

$$\left\{\sum_{\pi_0}\left[\sum_{j=1}^{n_i} x(\sigma_i^j)\alpha(\sigma_i^j)\right] - J(x, S)\right\} \subset \left(\sum_{0+\pi_0}\frac{1}{2^i}\right)U \subset 2U.$$

Let ν be a finite set of number pairs (i, j) which designate the superscript and subscript respectively of σ_j^i. Define $\nu' = [(i, j) \mid i \in \pi_0, j = 1, \cdots, n_i]$. If $\nu \geqq \nu'$, then the terms of $\sum_\nu x(\sigma_j^i)\alpha(\sigma_j^i)$ not included in the above sum may be split into two partial sums Σ' and Σ'' where Σ' contains only terms from $x(\sigma_i)$, $i \notin \pi_0$, and Σ'' contains only terms from $x(\sigma_i)$, $i \in \pi_0$. There exists π_1 ($\pi_1 \cdot \pi_0 = 0$) such that

$$\Sigma' \subset \sum_{i \notin \pi_0, i \in \pi_1} \operatorname{cvx}\,[x(\sigma_i)\alpha(\sigma_i) + \theta]$$

$$\subset \operatorname{cvx}\left[\sum_{\pi} x(\sigma_i)\alpha(\sigma_i) \,\middle|\, \text{finite set of } \pi\text{'s where } \pi \cdot \pi_0 = 0\right] \subset U.$$

Also

$$\Sigma'' \subset \sum_{i \in \pi_0, (i,j) \text{ of } \Sigma''} x(\sigma_{a_j}{}^i)\alpha(\sigma_{a_j}{}^i) \subset \sum_{\pi_0}\left[\sum_{(i,j) \text{ of } \Sigma''} x(\sigma_{a_j}{}^i)\alpha(\sigma_{a_j}{}^i)\right].$$

Now in the (i, j) of Σ'' we have $j > n_i$ so that $[\sum_{(i,j) \text{ of } \Sigma''} \alpha(\sigma_{a_j}{}^i)]x(\sigma_i) \subset U/2^i$.

By convexity, $\sum_{(i,\ j)\ \text{of}\ \Sigma''} x(\sigma_i)\alpha(\sigma_{a_j{}^i}) \subset U/2^i$, and hence $\Sigma'' \subset \sum_{\pi_0} U/2^i \subset U$. Finally if $\nu \geqq \nu'$,

$$\sum_{\nu} x(\sigma_j^i)\alpha(\sigma_j^i) - J(x, S) \subset 2U + U + U \subset 4U.$$

COROLLARY 2.4. *If $J_{\Delta_1}(x, \sigma)$ and $J_{\Delta_2}(x, \sigma)$ are unconditionally summable to $J(x, \sigma)$ with respect to U_1 and U_2 respectively, then $J_{\Delta_1 \cdot \Delta_2}(x, \sigma)$ is unconditionally summable to $J(x, \sigma)$ with respect to both $4U_1$ and $4U_2$.*

3. **Basic properties of the integral.** A set function $y(\sigma)$ on $\mathfrak{S}$ to X will be said to be *absolutely continuous* if for every $U \in \mathcal{U}$ there exists a $\delta_U > 0$ such that if $\alpha(\sigma) < \delta_U$ then $y(\sigma) \subset U$. Absolute continuity is clearly relative to $\alpha(\sigma)$.

THEOREM 3.1. *If $x(\sigma)$ is $\mathcal{U}$-integrable, then $\int_\sigma x d\alpha$ is absolutely continuous.*

Since $x(\sigma)$ is $\mathcal{U}$-integrable, it follows from Theorem 2.2 that to each $U \in \mathcal{U}$ there corresponds a Δ_U and π_U such that $\|\sum_{\pi U} x(\sigma \cdot \sigma_i)\alpha(\sigma \cdot \sigma_i) - J(x, \sigma)\|_U < 1/2$ for all $\sigma \in \mathfrak{S}$. As $\|x(\sigma_i)\|_U < \infty$ there exists a $\delta_U > 0$ for which if $\alpha(\sigma) < \delta_U$ then $\|x(\sigma_i)\alpha(\sigma)\|_U < 1/2^{i+1}$ $(i \in \pi_U)$. But then $\|\sum_{\pi U} x(\sigma \cdot \sigma_i)\alpha(\sigma \cdot \sigma_i)\|_U < 1/2$ and hence if only $\alpha(\sigma) < \delta_U$,

$$\|J(x, \sigma)\|_U \leqq \left\| \sum_{\pi U} x(\sigma \cdot \sigma_i)\alpha(\sigma \cdot \sigma_i) - J(x, \sigma) \right\|_U + \left\| \sum_{\pi U} x(\sigma \cdot \sigma_i)\alpha(\sigma \cdot \sigma_i) \right\|_U < 1.$$

THEOREM 3.2. *If $x(\sigma)$ is $\mathcal{U}$-integrable, then $\int_\sigma x d\alpha$ is additive.*

It follows from Theorem 2.2 that to each $U \in \mathcal{U}$ there corresponds a $\Delta_U = (\sigma_i)$ and π_U such that if $\pi \geqq \pi_U$, then $\|\sum_\pi x(\sigma \cdot \sigma_i)\alpha(\sigma \cdot \sigma_i) - J(x, \sigma)\|_U < 1/10$ for all $\sigma \in \mathfrak{S}$. Suppose $\sigma^0 = \sigma^1 + \sigma^2$ and $\sigma^1 \cdot \sigma^2 = 0$. The subdivision $(\sigma_i{}^0) = \{(\sigma^1 \cdot \sigma_j), (\sigma^2 \cdot \sigma_j)\}$ of σ^0 is greater than $(\geqq)$ the subdivision $(\sigma^0 \cdot \sigma_i)$ of σ^0. Hence by Theorem 2.3 there exists π_0 such that if $\pi \geqq \pi_0$ then

$$\left\| \sum_{\pi} x(\sigma_i{}^0)\alpha(\sigma_i{}^0) - J(x, \sigma^0) \right\|_U < 2/5.$$

When we combine these statements, taking $\pi \geqq \pi_0$ sufficiently large so that among the sets $\sigma_i{}^0$ $(i \in \pi)$ are the sets $(\sigma^1 \cdot \sigma_i)$ and $(\sigma^2 \cdot \sigma_i)$ for $i \in \pi_U$, we obtain

$$\begin{aligned} &\|J(x, \sigma^1) + J(x, \sigma^2) - J(x, \sigma^0)\|_U \\ &\quad \leqq \sum_{k=0}^{2} \left\| J(x, \sigma^k) - \sum_{\pi} x(\sigma^k \cdot \sigma_i{}^0)\alpha(\sigma^k \cdot \sigma_i{}^0) \right\|_U \\ &\quad + \sum_{k=1}^{2} \left\| \sum_{\pi} [x(\sigma^k \cdot \sigma_i{}^0)\alpha(\sigma^k \cdot \sigma_i{}^0) - x(\sigma^k \cdot \sigma_i{}^0)\alpha(\sigma^k \cdot \sigma_i{}^0)] \right\|_U < 1. \end{aligned}$$

Therefore $J(x, \sigma^0) = J(x, \sigma^1) + J(x, \sigma^2)$.

THEOREM 3.3. *If $x(\sigma)$ is $\mathcal{U}$-integrable, then $\int_\sigma x d\alpha$ is completely additive and $\sum_i \int_{\sigma_i} x d\alpha$ is unconditionally convergent.*

Let $\{\sigma_i\}$ be a sequence of disjoint sets of $\mathfrak{S}$. Since $\sum \alpha(\sigma_i) < \infty$, Theorems 3.1 and 3.2 imply $\lim_n \| \int_{\sum \sigma_i} x d\alpha - \sum_{i=1}^n \int_{\sigma_i} x d\alpha \|_U = 0$. As the same considerations hold for any subsequence of $\{\sigma_i\}$, the sum is by definition unconditionally convergent.

THEOREM 3.4. *If $x(\sigma)$ and $y(\sigma)$ are $\mathcal{U}$-integrable and m is a real number, then $m \cdot x(\sigma)$ and $x(\sigma)+y(\sigma)$ are $\mathcal{U}$-integrable, $mJ(x, \sigma) = J(mx, \sigma)$, and $J(x+y, \sigma) = J(x, \sigma) + J(y, \sigma)$.*

The conclusion about $m \cdot x(\sigma)$ follows from the fact that when $J_\Delta(x, \sigma)$ is unconditionally summable with respect to U to the value $J(x, \sigma)$, then $J_\Delta(mx, \sigma)$ is unconditionally summable to $mJ(x, \sigma)$ with respect to mU. Finally if $J_{\Delta_1}(x, \sigma)$ and $J_{\Delta_2}(x, \sigma)$ are unconditionally summable with respect to U to the values $J(x, \sigma)$ and $J(y, \sigma)$ respectively, then $J_\Delta(x+y, \sigma)$ where $\Delta \geqq \Delta_1 \cdot \Delta_2$ is unconditionally summable to $J(x, \sigma)+J(y, \sigma)$ with respect to $8U$ by Theorem 2.3.

4. **The integral in a complete space.** The conditions for integrability can be relaxed when X satisfies a certain type of completeness property. It is convenient to define this property by means of the general limit notion of E. H. Moore and H. L. Smith [10, p. 103] on a class D with a relation R on DD which is transitive and compositive. Each element of D will represent a unique subdivision Δ of S and neighborhood $U \in \mathcal{U}$, and will be designated by ΔU. We define $\Delta_1 U_1$ R $\Delta_2 U_2$ to mean $\Delta_1 \geqq \Delta_2$ and $U_1 \subset U_2$. A sequence $x_n \in X$ will be called a *fundamental sequence* if for each $U \in \mathcal{U}$ there is an N_U such that if $m, n \geqq N_U$ then $(x_n - x_m) \in U$. X will be said to be *sequentially complete* if every fundamental sequence is convergent. X will be said to be *complete with respect to D* if every "fundamental D-sequence," $x(\Delta U) \subset X$ (that is, every sequence such that for each $U \in \mathcal{U}$ there is a ΔU such that $\Delta_i U_i$ R ΔU $(i=1, 2)$ implies $[x(\Delta_1 U_1) - x(\Delta_2 U_2)] \in U$) is "convergent" (that is, an $x \in X$ exists such that for each $U \in \mathcal{U}$ there is a ΔU such that $\Delta_1 U_1$ R ΔU implies $[x(\Delta_1 U_1) - x] \in U$).

THEOREM 4.1. *If X is complete with respect to D and $x(\sigma)$ has the property that for every $U \in \mathcal{U}$ there exists a Δ_U such that $J_{\Delta_U}(x, S)$ is unconditionally summable with respect to U, then $x(\sigma)$ is $\mathcal{U}$-integrable.*

By hypothesis there exists Δ_U and π_U such that if $\pi_i \geqq \pi_U$ then

$$\left\| \sum_{\pi_1} x(\sigma_i)\alpha(\sigma_i) - \sum_{\pi_2} x(\sigma_i)\alpha(\sigma_i) \right\|_U < 1.$$

Let us designate by $\mathcal{J}_{\Delta U}(x, \sigma)$ some one element of the set $\sum_{\pi_U} x(\sigma\cdot\sigma_i)\alpha(\sigma\cdot\sigma_i)$. Theorem 2.2 implies that $\|\sum_{\pi} x(\sigma\cdot\sigma_i)\alpha(\sigma\cdot\sigma_i) - \mathcal{J}_{\Delta U}(x, \sigma)\|_U < 3$ for all $\sigma \in \mathfrak{S}$ if $\pi \geqq \pi_U$. By Theorem 2.3, if $\Delta_j = (\sigma_i{}^j) \geqq \Delta_U$, then there exists π_j such that for $\pi \geqq \pi_j$, $\|\sum_{\pi} x(\sigma\cdot\sigma_i{}^j)\alpha(\sigma\cdot\sigma_i{}^j) - \mathcal{J}_{\Delta U}(x, \sigma)\|_U < 12$. Finally when $\Delta_j U_j \text{ R } \Delta U$, we have

$$\left\| \sum_{\pi} x(\sigma\cdot\sigma_i{}^j)\alpha(\sigma\cdot\sigma_i{}^j) - \mathcal{J}_{\Delta_j U_j}(x, \sigma) \right\|_U \leqq \left\| \sum_{\pi} x(\sigma\cdot\sigma_i{}^j)\alpha(\sigma\cdot\sigma_i{}^j) - \mathcal{J}_{\Delta_j U_j}(x, \sigma) \right\|_{U_j} < 1$$

if only $\pi \geqq \pi_{U_j}$. Combining these statements gives the result that for $\Delta_j U_j \text{ R } \Delta U$ $(j = 1, 2)$,

$$\|\mathcal{J}_{\Delta_1 U_1}(x, \sigma) - \mathcal{J}_{\Delta_2 U_2}(x, \sigma)\|_U \leqq \sum_{j=1}^{2} \|\mathcal{J}_{\Delta_j U_j}(x, \sigma) + \mathcal{J}_{\Delta U}(x, \sigma)\|_U < 2\cdot 13 = 26.$$

Completeness with respect to D then implies the existence of $J(x, \sigma)$ for which if $U \in \mathcal{U}$ there exists a Δ_U and π_U such that if $\pi \geqq \pi_U$, then

$$\left\| J(x, \sigma) - \sum_{\pi} x(\sigma\cdot\sigma_i)\alpha(\sigma\cdot\sigma_i) \right\|_U < 1 + \|\mathcal{J}_{\Delta U}(x, \sigma) - J(x, \sigma)\|_U < 1 + 26 = 27.$$

LEMMA 4.2. *If X is sequentially complete and satisfies the first countability axiom, then X is complete with respect to D.*

Let $\{U_n\}$ where $U_n \supset U_{n+1}$ be a complete neighborhood system of θ. If $x(\Delta U)$ is a fundamental D-sequence, then for each $U \in \mathcal{U}$ there exists a $\Delta_U U \in D$ such that $\Delta_i U_i \text{ R } \Delta_U U$ $(i = 1, 2)$ implies $\|x(\Delta_1 U_1) - x(\Delta_2 U_2)\|_U < 1$. If we define $\Delta_n = \prod_{i=1}^{n} \Delta_{U_i}$, then $k_i \geqq n$ implies $\Delta_{k_i} U_{k_i} \text{ R } \Delta_{U_{k_i}} U_{k_i} \text{ R } \Delta_{U_n} U_n$ and $\|x(\Delta_{k_1} U_{k_1}) - x(\Delta_{k_2} U_{k_2})\|_{U_n} < 1$. As X is sequentially complete, there exists $x_0 = \lim_n x(\Delta_n U_n)$. Now given $U \in \mathcal{U}$, there exists $U_{n_0} \subset U$. Let $\Delta' = \Delta_U \cdot \Delta_{n_0}$. Then

$$\|x_0 - x(\Delta_U U)\|_U \leqq \|x_0 - x(\Delta_{n_0} U_{n_0})\|_U + \|x(\Delta_{n_0} U_{n_0}) - x(\Delta' U_{n_0})\|_{U_n} + \|x(\Delta' U_{n_0}) - x(\Delta_U U)\|_U \leqq 3.$$

COROLLARY 4.3. *If X is sequentially complete and satisfies the first countability axiom, and if $x(\sigma)$ has the property that for every $U \in \mathcal{U}$ there exists a Δ_U such that $J_{\Delta_U}(x, S)$ is unconditionally summable with respect to U, then $x(\sigma)$ is $\mathcal{U}$-integrable.*

This is an immediate consequence of Theorem 4.1 and Lemma 4.2.

5. **The continuous transform of an integrable function.** Let X and Y be linear topological spaces with the convex neighborhood systems $\mathcal{U}$ and $\mathcal{U}$ of their respective null elements. The following theorem can then be proved:

THEOREM 5.1. *If T is a linear continuous transformation on X to Y and if $x(\sigma)$ is $\mathcal{U}$-integrable, then $T[x(\sigma)]$ is $\mathcal{U}$-integrable and $J(T[x], \sigma) = T[J(x, \sigma)]$.*

Since T is continuous, for every $V \in \mathcal{U}$ there exists a $U \in \mathcal{U}$ such that $T(U) \subset V$. The $\mathcal{U}$-integrability of $x(\sigma)$ implies the existence of $J(x, \sigma)$, Δ_U, π_U such that if $\pi \geqq \pi_U$ then $\sum_\pi x(\sigma \cdot \sigma_i)\alpha(\sigma \cdot \sigma_i) - J(x, \sigma) \subset U$. Hence for $\pi \geqq \pi_U$ we have $\sum_\pi T[x(\sigma \cdot \sigma_i)]\alpha(\sigma \cdot \sigma_i) - T[J(x, \sigma)] \subset V$. It follows that $T[x(\sigma)]$ is $\mathcal{U}$-integrable and that $J(T[x], \sigma) = T[J(x, \sigma)]$.

COROLLARY 5.2. *If $\bar{x}$ is a linear continuous operation on X and if $x(\sigma)$ is $\mathcal{U}$-integrable, then $\bar{x}[x(\sigma)]$ is Lebesgue integrable and $\bar{x}[J(x, \sigma)] = \int_\sigma \bar{x}[x] d\alpha$.*

In this case Y is the space of real numbers. For real-valued functions our integral reduces to Fréchet's interpretation of the Lebesgue integral (11). The corollary then follows from Theorem 5.1.

If $x(s)$ is a multiple-valued point function on S to X and if the set function $x(\sigma) \equiv [x(s) \mid s \in \sigma]$ is $\mathcal{U}$-integrable to the value $J(x, \sigma)$, then we will say that $x(s)$ is $\mathcal{U}$-integrable to the value $J(x, \sigma)$. Further a set of linear continuous operations Γ on X will be said to be *total* if when $\bar{x}(x) = 0$ for all $\bar{x} \in \Gamma$, then $x = \theta$ [6, p. 42].

THEOREM 5.3. *If the set of linear continuous operations on X contains a denumerable subset which is total and if $x(s)$ and $y(s)$ are integrable to the same values on $\mathfrak{S}$, then $x(s) = y(s)$ a.e.*

Let $\{\bar{x}_n\}$ be a total denumerable subset of linear continuous operations on X. Then for all $\sigma \in \mathfrak{S}$, $\int_\sigma \bar{x}_n[x(s) - y(s)] d\alpha = 0$ by Corollary 5.2. Hence $\bar{x}_n[x(s) - y(s)] = 0$ on $S - \sigma_n$ where $\alpha(\sigma_n) = 0$. Therefore $\bar{x}_n[x(s) - y(s)] = 0$ on $S - \sigma_0 = S - \sum \sigma_i$ for all n, and $\alpha(\sigma_0) = 0$. As $\{\bar{x}_n\}$ is total, $x(s) = y(s)$ on $S - \sigma_0$, which proves the theorem.

The conjugate space $\overline{X}$ to a Banach space X has been defined by Dunford to be a fundamentally separable space with a determining manifold Γ in case Γ is a separable linear manifold in $\overline{X}$ such that for every $x \in X$, we have l.u.b. $[|\bar{x}(x)| \mid \bar{x} \in \Gamma, \|\bar{x}\| = 1] = \|x\|$ [2, p. 310]. It is clear that if X is a Banach space with its weak neighborhood topology or even its norm topology (see §1) and $\overline{X}$ is a fundamentally separable space, then the set of linear continuous operations on X contains a denumerable subset which is total. Theorem 5.3 is therefore applicable to such spaces.

6. **Convergence of integrable functions.** We prove first the following theorem.

THEOREM 6.1. *If $x_n(\sigma)$ are $\mathcal{U}$-integrable and if, for all $\sigma \in \mathfrak{S}$, $\{\int_\sigma x_n d\alpha\}$ is a fundamental sequence, then the integrals are equi-absolutely continuous.*

The argument of Saks (12) is applicable here. With the distance function $(\sigma_1, \sigma_2) = \alpha(\sigma_1 + \sigma_2 - \sigma_1 \cdot \sigma_2)$ the family of sets $\mathfrak{S}$ becomes a complete metric space. Let $F_n(\sigma) = \int_\sigma x_n d\alpha$ and $\mathfrak{S}_q \equiv [\sigma \in \mathfrak{S}$ such that $\|F_n(\sigma) - F_m(\sigma)\|_U \leqq e/3, m, n \geqq q]$. Since $F_n(\sigma)$ is absolutely continuous by Theorem 3.1, the sets $\mathfrak{S}_q$ are closed. Also $\mathfrak{S} = \sum_1^\infty \mathfrak{S}_q$. Hence by the Baire category theorem there exists an integer q_0 and a sphere $K(\sigma_0, r) \subset \mathfrak{S}_{q_0}$. Let $\delta < r$ be such that $\|F_{q_0}(\sigma)\|_U \leqq e/3$ if $\alpha(\sigma) < \delta$. Then for all σ such that $\alpha(\sigma) < \delta < r$, $\sigma_1 = \sigma + (\sigma_0 - \sigma)$ and $\sigma_2 = \sigma_0 - \sigma$ are in $K(\sigma_0, r)$ so that for $m, n \geqq q_0$,

$$\|F_n(\sigma) - F_m(\sigma)\|_U \leqq \|F_n(\sigma_1) - F_m(\sigma_1)\|_U + \|F_n(\sigma_2) - F_m(\sigma_2)\|_U \leqq 2e/3.$$

Therefore $\|F_n(\sigma)\|_U \leqq e$ if $\alpha(\sigma) < \delta$ and $n \geqq q_0$.

The integrable functions dealt with in the remainder of the paper will be single-valued point functions on S to X. As previously stated the integrability of $x(s)$ depends directly upon the integrability of the set function $x(\sigma) \equiv [x(s) \mid s \in \sigma]$, and the integral of $x(s)$ is defined to be that of $x(\sigma)$.

We would like to extend to X the notion of *approximate convergence* introduced by Dunford [2, p. 343]. Let $\sigma(n, e, U) \equiv [s \mid \|x_n(s) - x(s)\|_U \geqq e]$. The sequence $x_n(s)$ will be said to approach $x(s)$ approximately on S if for every n, U, and $e > 0$ there is a measurable set $\sigma'(n, e, U) \supset \sigma(n, e, U)$ such that $\lim_n \alpha[\sigma'(n, e, U)] = 0$.

We are now in a position to prove the following result:

THEOREM 6.2. *Let X be sequentially complete, let $x_n(s)$ be $\mathcal{U}$-integrable, and let $x_n(s) \to x(s)$ approximately. Then the following statements are equivalent:*

(1) *$x(s)$ is $\mathcal{U}$-integrable and*

$$\lim_n \int_\sigma x_n d\alpha = \int_\sigma x d\alpha$$

uniformly in σ.

(2) *$\lim_n \int_\sigma x_n d\alpha$ exists for every $\sigma \in \mathfrak{S}$.*

(3) *$\int_\sigma x_n d\alpha$ are equi-absolutely continuous.*

That (1) implies (2) is obvious, while Theorem 6.1 proves that (2) implies (3). There remains only the proof that (3) implies (1). For this purpose we demonstrate

LEMMA 6.3. *If $x(s)$ is $\mathcal{U}$-integrable and $\|x(s)\|_U \leqq e$, a.e., then $\|\int_\sigma x d\alpha\|_U \leqq e\alpha(\sigma)$.*

For given η and U there exist $\Delta_{\eta U}$ and $\pi_{\eta U}$ such that $\|\sum_{\pi_{\eta U}} x(\sigma \cdot \sigma_i)\alpha(\sigma \cdot \sigma_i) - J(x, \sigma)\|_U < \eta$ where $\Delta_{\eta U} = (\sigma_i)$ and $\sigma_1 \equiv [s \mid \|x(s)\|_U > e]$. But

$$\left\| \sum_{\pi_\eta U} x(\sigma\cdot\sigma_i)\alpha(\sigma\cdot\sigma_i) \right\|_U \leqq \sum_{\pi_\eta U} \|x(\sigma\cdot\sigma_i)\|_U \alpha(\sigma\cdot\sigma_i) \leqq e\alpha(\sigma).$$

Hence $\|J(x,\sigma)\|_U \leqq e\alpha(\sigma)+\eta$ where η is arbitrarily small.

To continue the proof of the theorem, we will first show that $\lim_n \int_\sigma x_n d\alpha$ exists uniformly in σ. By hypothesis, given U and $e>0$ there exists δ_{Ue} such that if $\alpha(\sigma)<\delta_{Ue}$, then $\|\int_\sigma x_n d\alpha\|_U < e$, and further, given U and $\delta>0$ there exists $N_{U\delta}$ such that if $n \geqq N_{U\delta}$ then $\alpha[\sigma'(n,\delta,U)]<\delta$. Let δ_0 be the smaller of e and δ_{Ue}. Then for m, $n \geqq N_{U\delta_0}$ we have $\|x_n(s)-x_m(s)\|_U < 2\delta_0$ on $S-\sigma'(n,\delta_0,U)-\sigma'(m,\delta_0,U)$. If $\sigma_{nm}=\sigma'(n,\delta_0,U)+\sigma'(m,\delta_0,U)$, then using Lemma 6.3,

$$\left\| \int_\sigma [x_n(s)-x_m(s)]d\alpha \right\|_U \leqq \left\| \int_{\sigma-\sigma_{nm}} [x_n(s)-x_m(s)]d\alpha \right\|_U + \left\| \int_{\sigma_{nm}} [x_n(s)-x_m(s)]d\alpha \right\|_U \leqq 2\delta_0\alpha(S)+2e \leqq (2\alpha(S)+2)e.$$

This together with the fact that X is sequentially complete implies that $\lim_n \int_\sigma x_n d\alpha$ exists uniformly in σ.

Given $U \in \mathcal{U}$, let us select a subsequence, which for notational convenience we will designate as $\{x_n(s)\}$, having the properties: (1) $\|\int_\sigma x_n d\alpha - \lim_n \int_\sigma x_n d\alpha\|_U < 1/2^n$ for all $\sigma \in \mathcal{S}$, and (2) $\|x_n(s)-x(s)\|_U < 1/(2^n\alpha(S))$ on $S-\tau_n$, where $\tau_n \in \mathcal{S}$, $\tau_n \supset \tau_{n+1}$, and $\alpha(\tau_n)<1/2^n$. Relation (2) is a consequence of approximate convergence. We will now show that $x(s)$ is $\mathcal{U}$-integrable to the value $J(\sigma)=\lim_n \int_\sigma x_n d\alpha$. As $x_n(s)$ is integrable, there exists a Δ_n and a π_n such that if $\pi \geqq \pi_n$ then

$$\left\| \sum_\pi x_n(\sigma\cdot\sigma_i^n)\alpha(\sigma\cdot\sigma_i^n) - J(x_n,\sigma) \right\|_U < \frac{1}{2^n}$$

where $\Delta_n=(\sigma_i^n)$. Hence for $\sigma \subset S-\tau_n$ and $\pi \geqq \pi_n$,

$$\begin{aligned}
\left\| \sum_\pi x(\sigma\cdot\sigma_i^n)\cdot\alpha(\sigma\cdot\sigma_i^n) - J(\sigma) \right\|_U &\leqq \text{l.u.b.}_{s_i \in \sigma\cdot\sigma_i^n} \left[\left\| \sum_\pi [x(s_i)-x_n(s_i)]\alpha(\sigma\cdot\sigma_i^n) \right\|_U \right. \\
&\quad \left. + \left\| \sum_\pi x_n(s_i)\alpha(\sigma\cdot\sigma_i^n) - J(x_n,\sigma) \right\|_U + \|J(x_n,\sigma)-J(\sigma)\|_U \right] \\
&< \text{l.u.b.}_{s_i \in \sigma\cdot\sigma_i^n} \left[\sum_\pi \|x(s_i)-x_n(s_i)\|_U \alpha(\sigma\cdot\sigma_i^n) + 1/2^n + 1/2^n \right] \\
&\leqq \frac{\alpha(\sigma)}{2^n\alpha(S)} + \frac{1}{2^{n-1}} \leqq \frac{1}{2^{n-2}}.
\end{aligned} \tag{1}$$

Statement (3) postulates for a given U, a $\delta_U>0$ such that if $\alpha(\sigma)<\delta_U$ then $\|\int_\sigma x_n d\alpha\|_U<1/2^3$ for all n. We now choose $n_0>5$ such that $\alpha(\tau_{n_0})<\delta_U$. Let $\bar\sigma_k{}^n=\sigma_k{}^n\cdot(\tau_{n-1}-\tau_n)$ for $n>n_0$ and let $\bar\Delta$ be the subdivision $(\bar\sigma_k{}^n)$ of τ_{n_0}. Then for any finite set n_i, k_i, $\|\sum_{i=1}^p x(\bar\sigma_{k_i}{}^{n_i})\alpha(\bar\sigma_{k_i}{}^{n_i})\|_U<1/2$. This can be demonstrated as follows: By the inequality (1), if $\pi\geqq\pi_n$, then

$$\left\|\sum_\pi\left[x(\bar\sigma_k{}^n)\alpha(\bar\sigma_k{}^n)-x(\bar\sigma_k{}^n)\alpha(\bar\sigma_k{}^n)\right]\right\|_U<\frac{1}{2^{n-3}}$$

and a fortiori

$$\left\|\sum_{n_i=n}\left[x(\bar\sigma_{k_i}^{n_i})\alpha(\bar\sigma_{k_i}^{n_i})-x(\bar\sigma_{k_i}^{n_i})\alpha(\bar\sigma_{k_i}^{n_i})\right]\right\|_U<\frac{1}{2^{n-3}}.$$

Therefore

$$\left\|\sum_{i=1}^p\left[x(\bar\sigma_{k_i}^{n_i})\alpha(\bar\sigma_{k_i}^{n_i})-x(\bar\sigma_{k_i}^{n_i})\alpha(\bar\sigma_{k_i}^{n_i})\right]\right\|_U<\frac{1}{2^{n_0-3}}\leqq\frac{1}{2^3}. \tag{2}$$

Let l be the greater of n_i $(i=1, 2, \cdots, p)$. Clearly there exists $\pi'\geqq\pi_l$ for which

$$\left\|\left[\sum_{t\,\varepsilon\,\pi}\alpha(\sigma_t\cdot\bar\sigma_{k_i}^{n_i})-\alpha(\bar\sigma_{k_i}^{n_i})\right]x(\bar\sigma_{k_i}^{n_i})\right\|_U<\frac{1}{2^3p}$$

whenever $\pi\geqq\pi'$. Hence if $\pi\geqq\pi'$,

$$\begin{aligned}
&\sum_{i=1}^p\left\{\left[\sum_{t\,\varepsilon\,\pi}x(\sigma_t{}^l\cdot\bar\sigma_{k_i}^{n_i})\alpha(\sigma_t{}^l\cdot\bar\sigma_{k_i}^{n_i})\right]-x(\bar\sigma_{k_i}^{n_i})\alpha(\bar\sigma_{k_i}^{n_i})\right\}\\
&\quad\subset\sum_{i=1}^p\left[\sum_{t\,\varepsilon\,\pi}x(\sigma_t{}^l\cdot\bar\sigma_{k_i}^{n_i})\alpha(\sigma_t{}^l\cdot\bar\sigma_{k_i}^{n_i})\right]-\sum_{i=1}^p\left[\sum_{t\,\varepsilon\,\pi}x(\bar\sigma_{k_i}^{n_i})\alpha(\sigma_t{}^l\cdot\bar\sigma_{k_i}^{n_i})\right]\\
&\qquad+\sum_{i=1}^p x(\bar\sigma_{k_i}^{n_i})\left[\sum_{t\,\varepsilon\,\pi}\alpha(\sigma_t{}^l\cdot\bar\sigma_{k_i}^{n_i})-\alpha(\bar\sigma_{k_i}^{n_i})\right]\\
&\quad=\sum_{i=1}^p\left\{\sum_{t\,\varepsilon\,\pi}\left[x(\sigma_t{}^l\cdot\bar\sigma_{k_i}^{n_i})-x(\bar\sigma_{k_i}^{n_i})\right]\alpha(\sigma_t{}^l\cdot\bar\sigma_{k_i}^{n_i})\right\}\\
&\qquad+\sum_{i=1}^p x(\bar\sigma_{k_i}^{n_i})\left[\sum_{t\,\varepsilon\,\pi}\alpha(\sigma_t{}^l\cdot\bar\sigma_{k_i}^{n_i})-\alpha(\bar\sigma_{k_i}^{n_i})\right]\\
&\quad\subset\sum_{i=1}^p\operatorname{cvx}\left[\left\{x(\bar\sigma_{k_i}^{n_i})-x(\bar\sigma_{k_i}^{n_i})\right\}\alpha(\bar\sigma_{k_i}^{n_i})\right]+\frac{1}{2^3}U\\
&\quad\subset\frac{1}{2^3}U+\frac{1}{2^3}U\subset\frac{1}{2^2}U.
\end{aligned}\tag{3}$$

The next to last inclusion follows because the inequality (2) holds for all subsets of $(1, 2, \cdots, p)$. Finally since the subdivision $\{(\sigma_j{}^l \cdot \bar{\sigma}_{k_i}{}^{n_i}), i=1, \cdots, p\}$ of $\sigma_0=\sum_{i=1}^{p}\bar{\sigma}_{k_i}{}^{n_i}$ is a repartition of the subdivision $(\sigma_0 \cdot \sigma_j{}^l)$, the inequality (1) and Theorem 2.3 imply the existence of $\pi'' \geqq \pi'$ such that

$$\left\| J(\sigma_0) - \sum_{i=1}^{p}\left\{ \sum_{l \in \pi''} x(\sigma_l{}^l \cdot \bar{\sigma}_{k_i}{}^{n_i})\alpha(\sigma_l{}^l \cdot \bar{\sigma}_{k_i}{}^{n_i}) \right\} \right\|_U < \frac{4}{2^{l-2}} \cdot \tag{4}$$

As $\|J(\sigma_0)\|_U < 1/2^3$, we have, after combining (3) and (4),

$$\left\| \sum_{i=1}^{p} x(\bar{\sigma}_{k_i}{}^{n_i})\alpha(\bar{\sigma}_{k_i}{}^{n_i}) \right\|_U < \frac{1}{2^2} + \frac{1}{2^{l-4}} + \frac{1}{2^3} \leqq \frac{1}{2} \cdot$$

We define a subdivision $\Delta_0=(\sigma_i)$ of S to be the sets of $\{\sigma_i{}^{n_0}(S-\tau_{n_0})\}$ and of $\bar{\Delta}$. Let π_0 be the π_{n_0} terms of $\{\sigma_i{}^{n_0} \cdot (S-\tau_{n_0})\}$. Then if $\pi \geqq \pi_0$, we have

$$\begin{aligned} \left\| \sum_{\pi} x(\sigma_i)\alpha(\sigma_i) - J(S) \right\|_U &\leqq \left\| \sum_{\text{sets of } \Delta_n} x(\sigma_i)\alpha(\sigma_i) - J(S-\tau_{n_0}) \right\|_U + \|J(\tau_{n_0})\|_U \\ &\quad + \left\| \sum_{\text{sets of } \bar{\Delta}} x(\sigma_i)\alpha(\sigma_i) \right\|_U \\ &\leqq \frac{1}{2^{n_0}} + \frac{1}{2^3} + \frac{1}{2} < 1. \end{aligned}$$

$x(s)$ is therefore integrable to $J(S)$. Clearly what has been shown for S could similarly be shown for any $\sigma \in \mathfrak{S}$.

We state the following theorem without proof:

THEOREM 6.4. *If $x_n(s)$ are $\mathcal{U}$-integrable $(n=0, 1, \cdots)$ and if $x_n(s) \to x_0(s)$ approximately, then the following are equivalent:*

(1) $\lim_n \int_\sigma x_n d\alpha = \int_\sigma x_0 d\alpha$ *exists uniformly in* σ.

(2) $\int_\sigma x_n d\alpha$ *are uniformly absolutely continuous.*

In the next theorem it is convenient to consider a pseudo-norm for the $\mathcal{U}$-integrable functions defined as $\|J(x)\|_U = \text{l.u.b.}_{\sigma \in S}\|J(x, \sigma)\|_U$.

THEOREM 6.5. *If X is sequentially complete, $x(s)$ $\mathcal{U}$-integrable, and $\phi(s)$ real-valued, measurable, and bounded by M, then $y(s)=\phi(s)x(s)$ is $\mathcal{U}$-integrable and $\|J(y)\|_U \leqq 2M\|J(x)\|_U$.*

Let $\phi_n(s)$ be a step function with the property that $|\phi_n(s)-\phi(s)| < 1/n$. As $x(s)$ is $\mathcal{U}$-integrable, there exists for a given $U \in \mathcal{U}$ a $\Delta_U=(\sigma_i)$ such that $J_{\Delta_U}(x, S)$ is unconditionally summable with respect to U. Hence for every k, $\|x(s)\|_U$ will be bounded on $\tau_k=\sum_{i=1}^{k}\sigma_i$ so that $\|(\phi_n(s)-\phi(s))x(s)\|_U < e$ for sufficiently large n on τ_k. Therefore $\phi_n(s)x(s) \to y(s)$ approximately. Suppose

$\phi_n(s) = c_i^n$ on σ_i^n. Then by a lemma due to Dunford [2, p. 313]

$$\left\| \int_\sigma \phi_n x_n d\alpha \right\|_U = \left\| \sum_1^{N_n} c_i^n \int_{\sigma \cdot \sigma_i^n} x d\alpha \right\|_U$$

$$(1) \qquad \leqq 2M \cdot \text{l.u.b.}_\pi \left[\left\| \sum_\pi \int_{\sigma \cdot \sigma_i^n} x d\alpha \right\|_U \Big| \ \pi \leqq (1, 2, \cdots, N_n) \right]$$

$$\leqq 2M \cdot \text{l.u.b.}_{\sigma^*} \left[\left\| \int_{\sigma \cdot \sigma^*} x d\alpha \right\|_U \Big| \ \sigma^* \in \mathfrak{S} \right].$$

The absolute continuity of $\int_\sigma x d\alpha$ (Theorem 3.1) thus implies the equi-absolute continuity of the integrals $\int_\sigma \phi_n x d\alpha$. By Theorem 6.2, $y(s)$ is $\mathcal{U}$-integrable and $\int_\sigma y d\alpha = \lim_n \int_\sigma \phi_n x d\alpha$. The remaining part of the theorem results immediately from (1).

7. **Instances of the $\mathcal{U}$-integral.** A further insight into the nature of the $\mathcal{U}$-integral can be gained by displaying some familiar integrals as representations of the $\mathcal{U}$-integral. We will first consider a very general theorem of this nature. To this end it will be recalled that in Theorem 1.1 we have shown that to every $U \in \mathcal{U}$ there corresponds a set of linear continuous operations and a range T_U such that $U \equiv [x \mid x \in X, \ |\bar{x}_t(x)| < 1$ for every $t \in T_U]$. Further, since the relation $\geqq$ on the set of subdivisions Δ is transitive and compositive, the general limit of E. H. Moore and H. L. Smith [10, p. 103] is applicable to this range.

THEOREM 7.1. *$x(s)$ is $\mathcal{U}$-integrable if and only if for every $U \in \mathcal{U}$, $\lim_\Delta \sum_{i=1}^\infty \bar{x}_t[x(\sigma \cdot \sigma_i)]\alpha(\sigma \cdot \sigma_i) = \bar{x}_t[J(x, \sigma)]$ uniformly in $t \in T_U$ and the $\sum_{i=1}^\infty \bar{x}_t[x(\sigma \cdot \sigma_i)]\alpha(\sigma \cdot \sigma_i)$ are absolutely convergent uniformly in $t \in T_U$.*

The statement that for every U and σ there exist $\Delta_{U\sigma}$ and $\pi_{U\sigma}$ such that if $\pi \geqq \pi_{U\sigma}$ then $\left\| \sum_\pi x(\sigma \cdot \sigma_i)\alpha(\sigma \cdot \sigma_i) - J(x, \sigma) \right\|_U < 1$ is equivalent by Theorems 1.1 and 2.3 to the statement that for every U and σ there exists $\Delta_{U\sigma}$ such that if $\Delta \geqq \Delta_{U\sigma}$ there is a π_Δ such that if $\pi \geqq \pi_\Delta$ then

$$\left| \sum_\pi \bar{x}_t[x(\sigma \cdot \sigma_i)]\alpha(\sigma \cdot \sigma_i) - \bar{x}_t[J(x, \sigma)] \right| < 1$$

for all $t \in T_U$.

COROLLARY 7.2. *If X is a Banach space with the weak neighborhood topology defined by all finite sets of elements in the adjoint space (that is, if $U \equiv [x \mid x \in X, \ |\bar{x}_t(x)| < 1, t = 1, \cdots, n]$), then the $\mathcal{U}$-integral is the integral defined by Pettis* [4, p. 280].

Each $\bar{x}$, element of the adjoint space, will clearly be continuous in the

weak topology and hence by Corollary 5.2 $\bar{x}[J(x, \sigma)] = \int_\sigma \bar{x}[x(s)]d\alpha$ for every $\sigma \in \mathfrak{S}$. Conversely, if the latter is true, then by Theorem 7.1, the $\mathcal{U}$-integral exists and is equal to the Pettis integral.

Dunford [2, p. 338] has considered a more general type of integral defined as follows: Let X be a closed linear manifold contained in $\overline{Y}$, the adjoint space to a Banach space Y. Then a function $y(s)$ on S to Y which has the property that $x[y(s)]$ is summable for every $x \in X$ defines uniquely an element in $\overline{X}$ according to the equation

$$\bar{x}_\sigma[x] = \int_\sigma x[y(s)]d\alpha.$$

$\bar{x}_\sigma$ is the Dunford integral of $y(s)$ over σ. However each $y \in Y$ corresponds to a unique element $\bar{x} \in \overline{X}$ through the correspondence $\bar{x}(x) = x(y)$ for all $x \in X$. The Dunford integral is unchanged when we replace the integrand $y(s)$ on S to Y by its correspondent $\bar{x}(s)$ on S to $\overline{X}$ (that is, $\int_\sigma x[y(s)]d\alpha = \int_\sigma \bar{x}(s)[x]d\alpha$ for all $x \in X$). If we make this replacement, we can prove by an argument similar to that used in Corollary 7.2 the following:

COROLLARY 7.3. *If $\overline{X}$ is adjoint to a Banach space X with the weak topology on $\overline{X}$ defined by all finite sets of elements of X, then the $\mathcal{U}$-integral of a function to $\overline{X}$ is the integral defined by Dunford.*

THEOREM 7.4. *If X is a Banach space and $\mathcal{U}$ its norm topology, then the $\mathcal{U}$-integral is the integral defined by Birkhoff.*

If $x(s)$ is Birkhoff integrable, then given $e > 0$ and $\sigma \in \mathfrak{S}$ there exists Δ_e and $J(x, \sigma)$ such that $J_{\Delta_e}(x, \sigma)$ is unconditionally convergent* to elements in an e-sphere about $J(x, \sigma)$. It is a fortiori unconditionally summable to $J(x, \sigma)$ with respect to an e-sphere and hence $\mathcal{U}$-integrable to $J(x, \sigma)$. Conversely, suppose $x(s)$ is $\mathcal{U}$-integrable. Then given $e/2^n$, there exists $\Delta_n = (\sigma_i^n)$, where $\Delta_{n+1} \geqq \Delta_n$, and π_n such that if $\pi \geqq \pi_n$ then $\left\|\sum_\pi x(\sigma_i^n)\alpha(\sigma_i^n) - J(x, S)\right\| < e/2^n$ $(n = 0, 1, \cdots)$. If N_n is the greater of the integers in π_n, then $m, i_k \geqq N_n$ implies that $\left\|\sum_1^m x(\sigma_i^n)\alpha(\sigma_i^n) - J(x, S)\right\| < e/2^n$ and $\left\|\sum_{k=1}^p x(\sigma_{i_k}^n)\alpha(\sigma_{i_k}^n)\right\| < 2e/2^n$ $(n = 0, 1, \cdots)$. Let P_n be the greater of N_n and M_n where $\sum_{i=M_n}^\infty \alpha(\sigma_i^n) < 1/2^n$. We define a sequence of sets $\sigma_1, \sigma_2, \cdots, \sigma_n, \cdots$ as follows: $\sigma_1 = \sigma_1^0$. Suppose σ_k is of Δ_n. Then σ_{k+1} is the set of lowest subscript of Δ_n which is disjoint from any of the previously chosen sets unless all such sets have subscripts greater than P_n. In the latter case, $R_n = k$ and σ_{k+1} is the set of lowest subscript of Δ_{n+1} which is disjoint from any of the previously chosen sets. $\sum_{i=1}^{R_n} \alpha(\sigma_i) \geqq \alpha(S) - 1/2^n$.

---

* If X is sequentially complete, then $\sum B_i$ where $B_i \subset X$ is said to be unconditionally convergent if given $U \in \mathcal{U}$ there exists N_U such that if $i_k \geqq N_U$ then $\left\|\sum_{k=1}^p B_{i_k}\right\|_U < 1$.

Hence $\Delta=(\sigma_i)$ is a subdivision of S. Now since $\Delta\geqq\Delta_0$, there exists by Theorem 2.3 an N_Δ such that $\|\sum_1^n x(\sigma_i)\alpha(\sigma_i)-J(x, S)\|<4e$ if only $n\geqq N_\Delta$. Finally by our choice of $\{\Delta_n\}$, if $i_k>R_n$ then σ_{ik} is a subset of some σ_i^n $(i\geqq N_n)$. Hence there exists π such that $\pi\cdot(1, \cdots, N_n)=0$ and

$$\left\|\sum_{k=1}^{p} x(\sigma_{i_k})\alpha(\sigma_{i_k})\right\| \leqq \left\|\sum_{\pi} \operatorname{cvx}\left[x(\sigma_i^n)\alpha(\sigma_i^n)+\theta\right]\right\|$$

$$\leqq \text{l.u.b.}\left[\left\|\sum_{\pi'} x(\sigma_i^n)\alpha(\sigma_i^n)\right\| \mid \pi'\leqq\pi\right]\leqq\frac{e}{2^{n-1}}.$$

$J_\Delta(x, S)$ is therefore unconditionally convergent to elements in an e-sphere about $J(x, S)$ and consequently $x(s)$ is Birkhoff integrable.

In §10 we demonstrate that the Birkhoff integral is distinct from the Pettis integral which in turn is distinct from the Dunford integral. Hence if we limit ourselves to Banach spaces on which we impose all possible convex neighborhood topologies which lie, say, in strength* between the norm topology and the usual weak topology, the corresponding ranges of integrable functions lie between the Birkhoff range and the Pettis range.

8. **Differentiation.** In the study of differentiation we wish to limit our attention to the subclass of $\mathcal{U}$-integrable functions on $S=(0, 1)$ to X defined as follows: $x(s)$ will be said to be $\mathcal{U}'$-*integrable* if

(1) $x(s)$ is $\mathcal{U}$-integrable.

(2) There exists a sequence of step functions† $\{x_n(s)\}$ such that $\|x_n(s)-x(s)\|_U\to 0$ a.e. for each $U\in\mathcal{U}$.

(3) $\|x(s)\|_U$ is summable for each $U\in\mathcal{U}$.

If X is a Banach space, then the $\mathcal{U}'$-integral is the Bochner integral [13]. On the other hand, if S is a Banach space with the usual weak neighborhood topology, then the $\mathcal{U}'$-integral is the Pettis integral [4]. This is easily seen from Corollary 7.2 and by writing $x_n(s)=\int_{I_i^n}x(s)ds/|I_i^n|$ if $s\in I_i^n$ where $I_i^n=((i-1)/n, i/n)$. Thus a peculiarity of this weak neighborhood topology is that the classes of $\mathcal{U}$- and $\mathcal{U}'$-integrable functions on $(0, 1)$ to X are the same (see Corollary 7.2).

We define a function $x(I)$ on intervals of $(0, 1)$ to X to be of *bounded variation*‡ if for every $U\in\mathcal{U}$ there exists an M_U such that for every finite set of disjoint intervals $I_1, \cdots, I_n$, $\sum_{i=1}^n\|x(I_i)\|_U\leqq M_U$. The function $x(I)$ will

* $\mathcal{U}$ is said to be weaker than $\mathcal{V}$ if whenever $x_n\to x$ in $\mathcal{V}$ then $x_n\to x$ in $\mathcal{U}$.

† A function on $(0, 1)$ to X will be called a *step function* if and only if it is single-valued on each of a finite number of disjoint intervals whose sum is $(0, 1)$.

‡ This is a generalization of the two notions of strong and weak bounded variation which have previously been introduced (see [3]).

be said to be *pseudo-differentiable** if there exists a $y(s)$ on $(0, 1)$ to X such that for every $U \in \mathcal{U}$, a.e.

$$\lim_{|I|\to 0} \left\| \frac{x(I)}{|I|} - y(s) \right\|_U = 0$$

where I is an arbitrary interval containing s. The function $y(s)$ is the *pseudo-derivative* of $x(I)$.

THEOREM 8.1. *A necessary and sufficient condition that a single-valued, additive, absolutely continuous function $x(\sigma)$ on measurable sets in $(0, 1)$ to X be pseudo-differentiable and of bounded variation is that the function $x(\sigma)$ be a $\mathcal{U}'$-integral. $x(\sigma)$ is then the $\mathcal{U}'$-integral of its pseudo-derivative.*

Proof of the necessity. Let $x_n(\ ) = x(I_i^n)/|I_i^n|$ on $I_i^n = (i/2^n, (i+1)/2^n)$ $(i=0, 1, \cdots, 2^n-1)$. Then by the definition of the pseudo-derivative there exists a $y(s)$ for which $\|x_n(s)-y(s)\|_U \to 0$ a.e. Since $\|x_n(s)\|_U$ is measurable and $\|x_n(s)\|_U \to \|y(s)\|_U$ a.e., it follows that $\|y(s)\|_U$ is measurable. As $x(\sigma)$ is of bounded variation, $\int_\sigma \|x_n(s)\|_U ds \leqq M_U$. Therefore by Fatou's lemma, $\|y(s)\|_U$ will be summable. It remains only to show that $y(s)$ is $\mathcal{U}$-integrable to the value $x(\sigma)$. Now given $e>0$ and $U \in \mathcal{U}$, there exists $\delta_{eU}>0$ such that if $|\sigma| < \delta_{eU}$ then $\int_\sigma \|y(s)\|_U ds \leqq e/8$. Further, since $\|x_n(s)-y(s)\|_U = \lim_{m\to\infty} \|x_n(s) - x_m(s)\|_U$ a.e. and is thus measurable, there exists σ_{eU} such that $\|x_n(s)-y(s)\|_U \to 0$ uniformly on $(0, 1)-\sigma_{eU} = \tau_{eU}$ and $|\sigma_{eU}| < \delta_{eU}$. Let n be chosen so that $\|x_n(s)-y(s)\|_U < e/4$ for $s \in \tau_{eU}$. If σ_i $(i=1, \cdots, N)$ are the sets on which $x_n(s)$ is constant in τ_{eU}, then

$$\left\| \sum_{i=1}^{N} y(\sigma\cdot\sigma_i)\cdot|\sigma\cdot\sigma_i| - \sum_{i=1}^{N} x_n(\sigma\cdot\sigma_i)\cdot|\sigma\cdot\sigma_i| \right\|_U < e/4.$$

Finally let σ_i $(i=N+1, \cdots)$ be a subdivision of σ_{eU} for which

$$\left\| \sum_{N+1}^{\infty} \|y(\sigma_i)\|_U |\sigma_i| - \int_{\sigma_{eU}} \|y(s)\|_U ds \right\| \leqq e/8.$$

Then for $i_k > N$,

$$\left\| \sum_{k=1}^{p} y(\sigma\cdot\sigma_{i_k})\cdot|\sigma\cdot\sigma_{i_k}| \right\|_U \leqq \sum_{k=1}^{p} \|y(\sigma\cdot\sigma_{i_k})\|_U \cdot |\sigma\cdot\sigma_{i_k}| \leqq \sum_{N+1}^{\infty} \|y(\sigma_i)\|_U |\sigma_i|$$

$$\leqq e/8 + \int_{\sigma_{eU}} \|y(s)\|_U ds \leqq e/4.$$

* Pettis has used this definition of pseudo-differentiability for a weak neighborhood topology on a Banach space.

For $\Delta=(\sigma_i)$ and $\pi_i \geqq (1, 2, \cdots, N)$ we have

$$\left\| \sum_{\pi_1} y(\sigma\cdot\sigma_i)\,|\,\sigma\cdot\sigma_i| - \sum_{\pi_2} y(\sigma\cdot\sigma_i)\,|\,\sigma\cdot\sigma_i| \right\|_U < e. \tag{1}$$

Now by Theorem 1.1 there exist linear continuous operations $\bar{x}_t$ on X such that $U \equiv [x \mid x \in X, |\bar{x}_t(x)| < 1 \text{ for every } t \in T_U]$. Our hypothesis clearly implies $\bar{x}_t[x(\sigma)] = \int_\sigma \bar{x}_t[y(s)]ds$.* There will therefore exist a subdivision $\Delta_t = (\sigma_i{}^t) \geqq \Delta$ for which

$$\left| \sum_i \bar{x}_t[y(\sigma\cdot\sigma_i{}^t)]\cdot|\,\sigma\cdot\sigma_i{}^t\,| - \bar{x}_t[x(\sigma)] \right| < e.$$

By (1) if $J_t(\sigma)$ is some element of $\sum_{\pi_0}\bar{x}_t[y(\sigma\cdot\sigma_i)]\,|\sigma\cdot\sigma_i|$ where $\pi_0 \equiv (1, 2, \cdots, N)$, then $\left|\bar{x}_t[\sum_\pi y(\sigma\cdot\sigma_i)\,|\sigma\cdot\sigma_i|\,] - J_t(\sigma)\right| < e$ for $\pi \geqq \pi_0$. Theorem 2.3 then implies

$$\left| \sum_i \bar{x}_t[y(\sigma\cdot\sigma_i{}^t)]\cdot|\,\sigma\cdot\sigma_i{}^t\,| - J_t(\sigma) \right| < 4e$$

and hence $|J_t(\sigma) - \bar{x}_t[x(\sigma)]| < 5e$. Therefore for $\pi \geqq \pi_0$,

$$\left| \bar{x}_t\left[\sum_\pi y(\sigma\cdot\sigma_i)\,|\,\sigma\cdot\sigma_i| - x(\sigma) \right] \right| < 6e$$

for every $t \in T_U$ or, what amounts to the same thing,

$$\left\| \sum_\pi y(\sigma\cdot\sigma_i)\cdot|\,\sigma\cdot\sigma_i| - x(\sigma) \right\|_U < 6e.$$

$y(s)$ is thus integrable to the value $x(\sigma)$.

Proof of the sufficiency. We will first prove a lemma.

LEMMA 8.2. *If $x(s)$ is $\mathcal{U}$-integrable and $\|x(s)\|_U$ is integrable for a given $U \in \mathcal{U}$, then $\|\int_\sigma x\,d\alpha\|_U \leqq \int_\sigma \|x(s)\|_U d\alpha$.*

Given $e>0$, there will be a Δ_e and π_e such that

$$\left\| \sum_{\pi_e} x(\sigma\cdot\sigma_i)\alpha(\sigma\cdot\sigma_i) - J(x, \sigma) \right\|_U < e/2,$$

$$\left| \sum_{\pi_e} \|x(\sigma\cdot\sigma_i)\|\alpha(\sigma\cdot\sigma_i) - \int_\sigma \|x(s)\|_U d\alpha \right| < e/2.$$

* If $x(\sigma)$ defined only on intervals is additive, absolutely continuous, of bounded variation, and has a pseudo-derivative, then this statement and the rest of the necessity remains valid for integrals on intervals. If X is in addition complete with respect to D, the integral can be extended on all measurable sets.

Since we have $\|\sum_{\sigma_i} x(\sigma\cdot\sigma_i)\alpha(\sigma\cdot\sigma_i)\|_U \leqq \sum_{\sigma_i}\|x(\sigma\cdot\sigma_i)\|_U\alpha(\sigma\cdot\sigma_i)$, it follows that $\|J(x,\sigma)\| \leqq \int_\sigma \|x(s)\|_U d\alpha + e$ where the e is arbitrary.

The proof of the sufficiency is now essentially an argument used by Bochner [13, pp. 269–270]. As $y(s)$ is $\mathcal{U}'$-integrable, there will exist a sequence $x_n^*(s)$ of step functions such that $\|x_n^*(s)-y(s)\|_U \to 0$ a.e. Define $x_n(s)=x_n^*(s)$ if $\|x_n^*(s)\|_U \leqq \|y(x)\|_U+1$, $x_n(s)=0$ otherwise. Then $\|x_n(s)-y(s)\|_U\to 0$ a.e.; $\|x_n(s)-y(s)\|_U$ is bounded by the summable function $(2\|y(s)\|_U+1)$; and $\|x_n(s)-y(s)\|_U$ is measurable since $\|x_n(s)-y(s)\|_U=\lim_m \|x_n(s)-x_m(s)\|_U$. Hence making use of the lemma,

$$\left\| \int_\sigma [x_n(s) - y(s)]ds \right\|_U \leqq \int_\sigma \|x_n(s) - y(s)\|_U ds.$$

Finally $\int_\sigma x_n(s)ds$ has a pseudo-derivative because $x_n(s)$ assumes only a finite set of values. Therefore

$$\begin{aligned}\lim_{|I|\to 0}\left\|\frac{\int_I y(s)ds}{|I|} - y(s)\right\|_U &\leqq \lim_{|I|\to 0}\frac{\int_I \|y(s) - x_n(s)\|_U ds}{|I|} \\ &\quad + \lim_{|I|\to 0}\left\|\frac{\int_I x_n(s)ds}{|I|} - x_n(s)\right\|_U + \|x_n(s) - y(s)\|_U \\ &\leqq 2\|x_n(s) - y(s)\|_U \text{ a.e.}\end{aligned}$$

As $\lim_n \|x_n(s)-y(s)\|_U=0$ a.e., $y(s)$ is the pseudo-derivative of $x(\sigma)$. Bounded variation follows from the lemma.

This theorem has previously been proved for the special cases where X is a Banach space [14, p. 410, footnote] and where X is a Banach space with its weak neighborhood topology [4, p. 300].

We may generalize the notion of derivative in another way: namely, a function $x(I)$ on intervals of $(0, 1)$ to X is said to be *differentiable* at a point s if $y(s) \in X$ exists such that for every U

$$\lim_{|I|\to 0}\left\|\frac{x(I)}{|I|} - y(s)\right\|_U = 0$$

where I is an arbitrary interval containing s. $y(s)$ is called the *derivative* of $x(I)$ at s. The vanishing of the limit is here independent of the choice of U. This kind of derivative does not seem to lend itself to a theorem analogous to Theorem 8.1. One should expect the integral of a well behaved integrand to have a derivative a.e. However in Example 10.9 we exhibit an additive, absolutely continuous function on measurable sets of $(0, 1)$ to a Hilbert space with its weak neighborhood topology which is pseudo-differentiable and of

bounded variation but not differentiable a.e. On the other hand its pseudo-derivative is both $\mathcal{U}'$-integrable and the limit in the norm sense of a sequence of step functions.

Pettis [23] has treated in detail the case $x(I)$ of bounded variation in the norm topology on a Banach space. In some special cases we have been able to demonstrate a few propositions about the derivative in a weak neighborhood topology. Theorem 5.1 of Pettis' paper is similar to Theorem 8.4 below.

In the remainder of this section X_1 will designate a Banach space and X the space X_1 with its weak neighborhood topology defined by all finite sets of linear limited operations on X_1.

$x(I)$ will be said to satisfy the *weak Lipschitz condition* (w.L.) if a.e. on $(0, 1)$ the norm of $x(I)/|I|$ is not infinite (that is, $\|x(I)/|I|\| < \infty$) for all intervals containing $s \in (0, 1)$.

THEOREM 8.3. *If $x(I)$ on intervals to X is differentiable a.e., then it satisfies w.L.*

As the neighborhoods of X are defined by all finite sets of elements of $\overline{X}_1$, adjoint to X_1, we have a.e. $\lim_{|I|\to 0} \bar{x}[x(I)/|I|]$ exists and a fortiori $|\bar{x}[x(I)/|I|]| < \infty$ for arbitrary $I \supset s$ and all $\bar{x} \in \overline{X}_1$. The theorem is then an immediate consequence of a theorem due to Dunford on uniform boundedness [2, p. 308, Theorem 2]. In this case the abstract set is the set of all intervals containing a point s at which the derivative exists.

THEOREM 8.4. *If X is sequentially complete with $\overline{X}$ separable or if X_1 is regular and if $x(I)$ on intervals to X is of bounded variation, then a necessary and sufficient condition that $x(I)$ be differentiable a.e. is that $x(I)$ satisfy w.L.*

The necessity has been proved in Theorem 8.3. For the sufficiency argument we first suppose $\overline{X}_1$ to be separable and obtain a sequence $\{\bar{x}_p\}$ dense in $\overline{X}_1$. Since $\bar{x}_p[x(I)]$ is of bounded variation, $\lim_{|I|\to 0} \bar{x}_p[x(I)]/|I|$ exists on a set σ_p of measure one. W.L. implies the existence of a function $L(s)$ such that $\bar{x}[x(I)]/|I| < L(s)\|\bar{x}\|$ on a set σ' of measure one. Therefore $\lim_{|I|\to 0} \bar{x}_p[x(I)]/|I|$ exists uniformly in p on a set $\sigma_0 = \sigma' \cdot \prod \sigma_p$ of measure one. It follows that $\lim_{|I|\to 0} \bar{x}[x(I)]/|I|$ exists on σ_0 for all $\bar{x} \in \overline{X}_1$. As X is sequentially complete, the derivative likewise exists on σ_0. If X_1 is regular and separable, then $\overline{X}_1$ is likewise separable. There remains only the case for which X_1 is a nonseparable regular Banach space. But here the above argument suffices because (1) $x(I)$ is contained in a separable closed linear subspace of X_1 and (2) every closed linear subspace of X_1 is regular [15, p. 423, Theorem 3]. (1) is a consequence of the following lemma:

LEMMA 8.5. *If $x(s)$ on $(0, 1)$ to a sequentially complete space X is of bounded variation, then $x(s)$ can have at most a denumerable number of points of discontinuity relative to the norm.*

Suppose the contrary to be true. Then there exists a positive number e and a denumerable set of points at which the norm variation of $x(s)$ is greater than e. We can then find a denumerable sequence of pairs of points with the property that $\|x(s_n^1)-x(s_n^2)\|>e$ and $s_n^1<s_n^2<s_{n^1+1}$ (or $s_n^1>s_n^2>s_{n^1+1}$). Let $y_n=x(s_n^1)-x(s_n^2)$. Since $x(s)$ is of bounded variation, $\sum_{n=1}^{\infty}|\bar{x}[y_n]|<\infty$ for every $\bar{x}\in\overline{X}_1$. The sequential completeness of X implies that every subseries of $\sum_n y_n$ is convergent in X. By a theorem due to Orlicz and Banach [4, pp. 281–282, Theorem 2.32], $\sum y_n$ is unconditionally convergent in X_1 and hence $\|y_n\|\to 0$, which is contrary to our choice of y_n's.

Returning to the theorem, it is clear from the lemma that $x(s)=x(I_0^s)$ has only a denumerable set of points of discontinuity relative to the norm and hence all of its values can be found as limits of a denumerable set of values (again in the sense of the norm). $x(I)$ is therefore contained in the separable linear closed subspace of X_1 determined by this denumerable set of values of $x(s)$. This completes the proof of Theorem 8.4.

COROLLARY 8.6. *If X is sequentially complete with $\overline{X}_1$ separable or if X_1 is regular, and if $x(I)$ is additive, absolutely continuous, of bounded variation, and of w.L., then $x(I)$ is an integral.*

This follows from Theorem 8.4 and a theorem due to Pettis [4, pp. 300–301].

In the final theorem of this section we suppose X_1 to be a Banach space with a base $\{x_i\}$ [6, p. 110] which satisfies the following postulate:

(M) If $a_1, a_2, \cdots$ is any sequence of real numbers such that l.u.b.$_n$ $\|\sum_1^n a_i x_i\|<\infty$, then $\sum_1^{\infty} a_i x_i$ converges.

By redefining the norm we can consider instead of X_1 a space isomorphic [6, p. 180] to X_1 satisfying in addition to (M) the property

$$\text{(N)}\qquad \left\|\sum_1^n a_i x_i\right\| \leqq \left\|\sum_1^{n+1} a_i x_i\right\|$$

for any constants $a_1, a_2, \cdots$ [16, pp. 415–416]. The spaces l_p $(p\geqq 1)$ and L_p $(p>1)$ satisfy (M) [16, p. 419].

THEOREM 8.7. *If $x(I)$ is an additive, absolutely continuous function satisfying w.L. on intervals of $(0, 1)$ to X where X_1 satisfies* (M) *and* (N), *then $x(I)$ is pseudo-differentiable and is the integral of its pseudo-derivative on intervals.*

Now $x(I)=\sum_1^{\infty} a_i(I)x_i$ where $a_i(I)=\bar{x}_i[x(I)]$ and hence are additive abso-

lutely continuous functions of intervals with summable derivatives $a_i'(s)$ a.e. As $x(I)$ satisfies w.L.,

$$\left\| \sum_{i=1}^{n} \frac{a_i(I)}{|I|} x_i \right\| \leqq \left\| \frac{x(I)}{|I|} \right\| \leqq L_s \text{ a.e.}$$

so that $\|\sum_1^\infty a_i'(s)x_i\| \leqq L_s$ a.e. By property (M) $y(s) = \sum_1^\infty a_i'(s)x_i$ exists a.e. Let $x_n(I) = \sum_1^n [\int_I a_i'(s)ds]x_i$. Then $\|x(I)\| \geqq \|x_n(I)\|$ implies that the $x_n(I)$ are uniformly absolutely continuous. Hence

$$\bar{x}[x(I)] = \lim_n \bar{x}[x_n(I)] = \lim_n \int_I \left[\sum_1^n a_i'(s)\bar{x}(x_i) \right] ds$$

$$= \int_I \left[\sum_1^\infty a_i'(s)\bar{x}(x_i) \right] ds = \int_I \bar{x}[y(s)]ds.$$

This is equivalent to the desired conclusions.

9. **An application to differential equations.** We consider in this section an existence theorem for ordinary differential equations in X similar to a real variable theorem due to Carathéodory [17, p. 672].* We limit ourselves to linear convex topological spaces which are sequentially complete and satisfy the first countability axiom. Garrett Birkhoff [17] has shown that such a space has a group metric. X is therefore an F-space [6, p. 34] with a convex neighborhood topology. We will describe the topology by both a metric and the convex neighborhoods as suits our convenience.

$x(s)$ and $\phi(x, s)$ will be functions on $(0, 1)$ and the product space $X \times (0, 1)$ respectively to X. $Dx(s)$ will designate the pseudo-derivative of $x(s)$. Since X satisfies the first countability axiom, the existence of the pseudo-derivative will imply the existence of the derivative a.e. (see §8). $\int_s x(s)ds$ will be the $\mathcal{U}'$-integral. The system

$$Dx = \phi(x, s) \qquad x(0) = x_0 \tag{1}$$

will be said to have a *solution* $x(s)$ if and only if

$$x(s) = x_0 + \int_0^s \phi[x(t), t]dt.$$

$x(s)$ will be called *measurable* if there exists a sequence of step functions $\{x_n(s)\}$ such that $|x_n(s) - x(s)| \to 0$ a.e. where $|x-y|$ is the distance between x and y. We will say that the sequence of functions $\{x_n(s)\}$ are *equally-continuous* if for every $U \in \mathcal{U}$ there exists $\delta_U > 0$ such that $|s_1 - s_2| < \delta_U$ implies

* The basic ideas of this theorem were also in part suggested by T. H. Hildebrandt.

$\|x(s_1)-x(s_2)\|_U<1$. Finally $\phi(x, s)$ will be said to satisfy *condition* (A) if it is continuous in x for fixed s and measurable in s for fixed x.

THEOREM 9.1. *A necessary and sufficient condition that $x(s)$ be a solution of* (1) *is that* (a) $x(0)=x_0$, (b) $x(I_{s_1}^{s_2})=x(s_2)-x(s_1)$ *be an additive, absolutely continuous function of bounded variation on intervals of* (0, 1) *to X, and* (c) $Dx(s)$ *exist and be equal to* $\phi[x(s), s]$.

This is an immediate consequence of Theorem 8.1 and the footnote on page 133.

LEMMA 9.2. *If $\phi(x, s)$ satisfies* (A) *and $x(s)$ is measurable on* (0, 1) *to X, then $\phi[x(s), s]$ is measurable.*

Since $x(s)$ is measurable, there exists a sequence of step functions $\{x_n(s)\}$ such that $|x_n(s)-x(s)|\to 0$ a.e. and by the continuity of $\phi(x, s)$ in x, $|\phi[x_n(s), s]-\phi[x(s), s]|\to 0$ a.e. Now $\phi[x_n(s), s]$ is clearly measurable. There will therefore exist a double sequence of step functions $\{y_n^p(s)\}$ such that for every n, $\lim_p |y_n^p(s)-\phi[x_n(s), s]|=0$ a.e. As the real-valued functions $|y_n^p(s)-\phi[x_n(s), s]|$ are measurable, there will exist the set $y_n^{p_n}$ for which $|y_n^{p_n}-\phi[x_n(s), s]|<1/2^n$ except on a set of measure less than $1/2^n$. Therefore $|y_n^{p_n}-\phi[x(s), s]|\to 0$, a.e.

LEMMA 9.3. *If $\{x_n(s)\}$ on* (0, 1) *to X are equally continuous and compact-valued for each s ε (0, 1), then there exists a subsequence $\{x_{n_i}(s)\}$ which converges uniformly to a continuous function on* (0, 1) *to X.*

The proof is similar to that of the familiar Arzelas theorem and will be left to the reader.

LEMMA 9.4. *If Γ is a compact subset of X and $M>0$, then the set of all elements of the type $\sum_1^n a_i x_i$ where $x_i \varepsilon \Gamma$ and $\sum_1^n |a_i| \leqq M$ form a compact set.*

Let $\{U_n\}$ where $U_n \supset U_{n+1}$ be a complete neighborhood system for θ and let $e=1/2(M+1)$. There will exist a sequence $\{\eta_i\} \subset \Gamma$ and integers N_n such that for $x \varepsilon \Gamma$ there is an η_i ($i \leqq N_n$) such that $\|x-\eta_i\|_{U_n}<e$. In the nth step we let correspond to every $x \varepsilon \Gamma$ an η_i ($i \leqq N_n$ and $\|x-\eta_i\|_{U_n}<e$) and then to every $\sum_1^p a_j x_j$ the element $\sum_1^{N_n} A_{ni}\eta_i$ where $A_i=\sum a_j$ summed over all j such that x_j corresponds to η_i. Let $\{\sum_1^{p_r} a_j^r x_j^r\}$ be a sequence of such elements and let $\{\sum_1^{N_n} A_{ni}^r \eta_i\}$ be their correspondents. Then there exists a subsequence $\{r_k\}$ for which $A_{ni}^r \to \overline{A}_i^n$ ($i=1, \cdots, N_n$). Hence there will be an N_n' such that $k \geqq N_n'$ implies

$$\left\| \sum_1^{N_n} A_{ni}^{r_k}\eta_i - \sum_1^{N_n} \overline{A}_i^n \eta_i \right\|_{U_n} \leqq \sum_1^{N_n} |A_{ni}^{r_k} - \overline{A}_i^n| \cdot \|\eta_i\|_{U_n} < e.$$

Clearly we can apply the diagonal process and obtain a subsequence $\{s_k\}$ for which $\|\sum_1^{N_n} A_{ni}^{s_k}\eta_i - \sum_1^{N_n}\overline{A}_i{}^n\eta_i\|_{U_n} < e$ if $k \geqq N_n''$. Now

$$\left\| \sum_1^{p_{s_k}} a_i^{s_k} x_i^{s_k} - \sum_1^{N_n} A_{ni}^{s_k}\eta_i \right\|_{U_n} < Me.$$

Therefore

$$\left\| \sum_1^{p_{s_k}} a_i^{s_k} x_i^{s_k} - \sum_1^{p_{s_l}} a_i^{s_l} x_i^{s_l} \right\|_{U_n} < 2(M+1)e$$

if only $k, l \geqq N_n''$. As X is sequentially complete, the sequence $\{\sum^{p_{s_k}} a_i^{s_k} x_i^{s_k}\}$ converges to an element of X.

We are now in a position to establish the existence theorem.

THEOREM 9.5. *If $\phi(x, s)$ satisfies* (A) *and takes on only a compact set of values for $s \in (0, 1)$ and $x \in X$, then the system* (1) *has a solution.*

We first show that for any measurable function $x(s)$, $\phi[x(s), s]$ can be uniformly approximated by simple functions* assuming values in $\Gamma \equiv [\phi(x, s) \mid s \in (0, 1),\ x \in X]$. Let $\{U_n\}$ where $U_n \supset U_{n+1}$ be a complete neighborhood system for θ. There will exist a sequence $\{\eta_i\} \subset \Gamma$ and integers N_n such that for $x \in \Gamma$ there is an η_i $(i \leqq N_n)$ for which $\|x - \eta_i\|_{U_k} \leqq 1$ $(k \leqq n)$. We now define $S_n{}^i \equiv [s \mid s \in (0, 1),\ \|\eta_i - \phi[x(s), s]\|_{U_n} \leqq 1]$, $\sigma_n{}^i = S_n{}^i - \sum_{j=1}^{i-1} S_n{}^j$ $(i = 1, \cdots, N_n)$, and $y_n(s) = \eta_i$ on $\sigma_n{}^i$. To show that $\{y_n(s)\}$ is the desired sequence we need only to demonstrate that $S_n{}^i$ is measurable. But this follows from the fact that $\phi[x(s), s]$ is approximable by step functions.

By hypothesis, Γ is compact. Therefore given $U \in \mathcal{U}$, there will exist $M(U)$ such that $\|\Gamma\|_U \leqq M(U)$. Further if $x(s)$ is measurable, then $\phi[x(s), s]$ is measurable, $\|\phi[x(s), s]\|_U$ is summable, and hence the $\mathcal{U}$-integrability of $\phi[x(s), s]$ will imply its $\mathcal{U}'$-integrability.

Let $x_n(s) = x_0 + \int_0^s \phi[x_n'(s), s]ds$ where $x_n'(s) = x(p/n)$ for $p/n < s \leqq (p+1)/n$ $(p = 0, 1, \cdots, n-1)$. The integral of $\phi[x_n'(s), s]$ exists by Theorem 6.2, since $\phi[x_n'(s), s]$ is uniformly approximable by simple functions whose pseudo-norms with respect to U are bounded by $M(U)$. Now by Lemma 8.2

$$\|x_n(s_1) - x_n(s_2)\|_U \leqq \left\| \int_{s_1}^{s_2} \phi[x_n'(s), s]ds \right\|_U \leqq \int_{s_1}^{s_2} \|\phi[x_n'(s), s]\|_U ds \leqq M(U) \cdot |s_2 - s_1|. \tag{2}$$

This clearly implies

$$\|x_n(s) - x_n'(s)\|_U \leqq M(U)/n. \tag{3}$$

* A simple function on (0, 1) to X is constant on each of a finite number of disjoint measurable sets whose sum is (0, 1).

As above there exists a double sequence of simple functions $y_p^n(s)$ such that $y_p^n(s)=\eta_i$ on $\sigma_{pi}^n \in \mathcal{S}$ ($\eta_i \in \Gamma$ and $\sum_{i=1}^{N_p}\sigma_{pi}^n=S$) and $\|y_p^n(s)-\phi[x_n'(s), s]\|_{U_p}\leqq 1$. It is clear that $\|y_p^n(s)-\phi[x_n'(s), s]\|_{U_k}$ is measurable and converges to zero as $p\to\infty$. Thus $\|y_p^n(s)\|_U \leqq M(U)$ and $y_p^n(s)\to\phi[x_n'(s), s]$ approximately. By Theorem 6.2 $\lim_p \int_\sigma y_p^n(s)ds=\int_\sigma \phi[x_n'(s), s]ds$. But $\int_\sigma y_p^n(s)ds=\sum_1^{N_p}\eta_i|\sigma\cdot\sigma_{pi}^n|$ where $\sum_{i=1}^{N_p}|\sigma\cdot\sigma_{pi}^n|\leqq 1$. By Lemma 9.4 the set $[\int_\sigma y_p^n(s)ds \mid \sigma \subset (0, 1); n; p=1, 2, \cdots]$ is compact and therefore $[\int_0^s \phi[x_n'(s), s]ds \mid s \in (0, 1); n=1, 2, \cdots]$ is compact. By (2) and Lemma 9.3 there exists a subsequence $\{x_{n_k}(s)\}$ which approaches the continuous function $x(s)$ uniformly on $(0, 1)$. Then by (3) $x_{n_k}'(s)\to x(s)$ uniformly on $(0, 1)$. Clearly $\|\phi[x_{n_k}'(s), s]-\phi[x(s), s)]\|_U$ is measurable. Therefore $\phi[x_{n_k}'(s), s]\to\phi[x(s), s]$ approximately as $k\to\infty$. Also $\|\phi[x_{n_k}'(s), s]\|_U\leqq M(U)$. Again by Theorem 6.2

$$x(s)-x_0=\lim_k \int_0^s \phi[x_{n_k}'(s), s]ds=\int_0^s \phi[x(s), s]ds.$$

10. **Examples.** The following examples serve to illustrate some of the properties of the $\mathcal{U}$-integral as well as the importance of various hypotheses. Example 10.2 demonstrates that the Birkhoff and Pettis integrals are different. It is therefore possible to define a class of integrals for a Banach space. Example 10.8 shows that the class of Pettis integrable functions does not include a very well behaved set of functions. Example 10.9 proves that the integral of a strongly measurable function may not be weakly differentiable. These three examples answer questions raised by Pettis [4, p. 303].

B will be the space of real bounded functions x_t on $T=(0, 1)$ with norm $\|x_t\|=\text{l.u.b.}_{t\in T}|x_t|$. We will consider functions $x(s)$ on $S=(0, 1)$ to X. Integration on $S=(0, 1)$ is taken with respect to the Lebesgue measure function.

It will be necessary for us to refer to the following theorem whose proof can be found in another paper* by the author:

THEOREM 10.1. *For every bounded additive set function $\beta(\tau)$ on all subsets of $(0, 1)$ there exists a denumerable set $R(\beta)$ and a unique decomposition $\beta(\tau)=\beta_1(\tau)+\beta_2(\tau)$ such that $\beta_i(\tau)$ $(i=1, 2)$ are bounded additive set functions, $\beta_1(\tau)=\beta(R\cdot\tau)$, and $\beta_2(\tau)=0$ if τ is a denumerable subset of $(0, 1)$.*

10.2. *Example of a function which is Pettis integrable but not Birkhoff integrable* (§7). Let $x(s)$ on $S=(0, 1)$ to B be defined as $x(s)\equiv x_t(s)=1$ if $t-s=p/2^n$ for $\pm p$, n positive integers and $x_t(s)=0$ elsewhere. Thus on the unit square with coordinates s and t, $x_t(s)=1$ only on a denumerable set of points on any line parallel to either coordinate axis.

---

* This paper will appear in the Bulletin of the American Mathematical Society.

Hildebrandt [19] has shown that for every $\bar{x}$ of the space adjoint to B there exists an additive bounded set function $\beta(\tau)$ on all subsets of $T=(0, 1)$ such that for all $x \in B$, $\bar{x}(x)=\int_T x_t d\beta$. By Theorem 10.1, $\beta(\tau)=\beta_1(\tau)+\beta_2(\tau)$, where $\beta_1(\tau)=\beta(R\cdot\tau)$, R is a denumerable subset of T, and $\beta_2(\tau)=0$ for all denumerable subsets of T. Hence for each s

$$\bar{x}[x(s)] = \int_T x_t(s)d\beta = \int_T x_t(s)d\beta_1 + \int_T x_t(s)d\beta_2 = \int_T x_t(s)d\beta_1.$$

It follows that $\bar{x}[x(s)]\neq 0$ only if $x_t(s)\neq 0$ for $t \in R$. Let $\delta_t \equiv [s \mid s \in (0, 1), x_t(s)\neq 0, \text{fixed } t]$ and $\delta_0=\sum_{t\in R}\delta_t$. Then δ_0 is the subset of S on which $x_t(s)\neq 0$ for some $t \in R$. Since both δ_t and R are denumerable sets, δ_0 is a denumerable subset of S. Therefore $\bar{x}[x(s)]=0$ a.e. and $x(s)$ is Pettis integrable to the value $x_\sigma=\theta \in B$ for all measurable sets σ.

On the other hand $x(s)$ is not Birkhoff integrable. Suppose the contrary were true. Then [1, Theorem 13], given $e>0$, there exist disjoint sets σ_i such that $\sum\sigma_i=S$, the series $\sum x(\sigma_i)\alpha(\sigma_i)$ is unconditionally convergent, and $s_i, s_i' \in \sigma_i$ implies $\|\sum_{i=1}^{\infty} x(s_i)|\sigma_i| - \sum_{i=1}^{\infty} x(s_i')|\sigma_i|\| < e$. We first suppose that for each t there exists a σ_i such that for all $s \in \sigma_i$, $x_t(s)=0$. We can then divide T into sets θ_i where $T=\sum\theta_i$, and for all $t \in \theta_i$ and $s \in \sigma_i$, $x_t(s)=0$. Since $x_t(s)=1$ on a diagonal, $S-\sigma_i$ contains the set θ_i' whose coordinate values are those of θ_i. In addition $S-\sigma_i$ contains θ_i'' consisting of all points of the type $s\pm p/2^n$ and $s/2^n$ where $s \in \theta_i'$ for all integral p and n. Now θ_i'' is a homogeneous point set in the sense of Knopp [20, pp. 411–414] so that its upper measure $|\overline{\theta_i''}|=1$ or 0. As there exists an i_0 such that $|\overline{\theta_{i_0}'}|>0$, it follows that $|S-\sigma_{i_0}| \geq |\overline{\theta_{i_0}''}|=1$ which is contrary to $|\sigma_{i_0}|>0$. Our supposition being false, we can select a t_0 with the property that for every i, σ_i contains an s_i such that $x_{t_0}(s_i)=1$. As $|\sigma_i|>0$, there is also an $s_i' \in \sigma_i$ such that $x_{t_0}(s_i')=0$. But this implies that $\|\sum_i x(s_i)|\sigma_i| - \sum_i x(s_i')|\sigma_i|\| \geq 1$. The assumption that $x(s)$ is Birkhoff integrable has thus led to a contradiction.

10.3. *Example of a $\mathcal{V}$-integrable function for which there exists Δ_0 such that for all $\Delta \geq \Delta_0$, $J_\Delta(x, S)$ is not unconditionally convergent* (§2). One would expect a definition of unconditional convergence for $J_\Delta(x, S)$ at least to demand the convergence of every sum $\sum_1^{\infty} x(s_i)|\sigma_i|$ $(s_i \in \sigma_i)$. Even such a definition would eliminate the function of the present example from the class of integrable functions if (1) we replace unconditional summability in Definition 2.1 by unconditional convergence, and (2) we insist that $J_\Delta(x, S)$ is unconditionally convergent if this is true for $J_{\Delta_0}(x, S)$ where $\Delta \geq \Delta_0$.

Let X be the space B with a weak neighborhood topology defined by all finite sets of linear limited operations on B. Let $x(s)\equiv x_t(s)=1/s$ for $s-t=p/2^n$ where $\pm p$, n are positive integers and $x_t(s)=0$ elsewhere. $x(s)$ can be shown

to be $\mathcal{U}$-integrable (that is, Pettis integrable) in a manner identical with that used in Example 10.2. Suppose $\Delta_0 = \{\sigma_i^0 = (1/2^i, 1/2^{i-1})\}$ and $\Delta = (\sigma_i) \geqq \Delta_0$. As in Example 10.2 there exists t_0 and $s_i \in \sigma_i$ such that $x_{t_0}(s_i) = 1/s_i$. We will now show that some sum of $J_\Delta(x, S)$ does not converge with respect to the neighborhoods $U \equiv [x \mid x \in X, \ |\hat{x}_{t_0}(x)| < \delta]$ where $\hat{x}_{t_0}(x) = x_{t_0}$. This is an immediate consequence of

$$\sum_i x_{t_0}(s_i) \, |\, \sigma_i \,| \geqq \sum_i \frac{1}{1/2^{i-1}} |\, \sigma_i^0 \,| = \sum_i \frac{2^{i-1}}{2^i} = \infty .$$

10.4. *Example of a function $x(s)$ on $S = (0, 1)$ to B which is $\mathcal{U}$-integrable having the properties* (1), (2), (3) *given below.*

(1) $\|x(s)\|$ *is not measurable.* Convergence a.e. will therefore not imply uniform convergence on the complement of sets of arbitrary small measure (approximate convergence in Theorem 6.2).

(2) $\int_\sigma x(s)ds = \theta$ *for all σ yet $x(s) \neq 0$ a.e.* $\overline{B}$ does not contain a denumerable subset which is total (Theorem 5.2).

(3) $x(s)$ *is not $\mathcal{U}'$-integrable.* The two integrals are therefore distinct in general (§8).

Let σ_0 be a non-measurable set in $S = (0, 1)$ and define $x(s) = \theta$ for $s \in S - \sigma_0$ and $x(s)$ equal to the characteristic function of the point s for $s \in \sigma_0$ [4, p. 301, Example 9.1].

10.5. *Example illustrating the need for sequential completeness in Corollary* 4.3 *and Theorem* 6.2. In the Banach space of bounded sequences let m_0 be the linear extension of the set of unit vectors ϕ_n. The space m_0 is not sequentially complete. Let $x^k(s) \equiv x_i^k(s) = \delta_{in}\phi_n$ for $s \in (1/2^n, 1/2^{n-1})$ and $n \leqq k$ (δ_{in} is the Kronecker symbol), and vanish elsewhere ($k = 1, 2, \cdots, \omega$). Despite the fact that for every $e > 0$ there exists a $J_\Delta(x^\omega, S)$ which is unconditionally summable with respect to an e-sphere, $x^\omega(s)$ is not $\mathcal{U}$-integrable in m_0 (Corollary 4.3). On the other hand $x^k(s) \to x^\omega(s)$ approximately, and the $\int_\sigma x^k(s)ds$ exist and are uniformly absolutely continuous (Theorem 6.2).

10.6. *Example showing the need for sequential completeness in Theorem* 6.5. Let B_0 be the linear subset of B consisting of all functions on $T = (0, 1)$ whose derivatives exist a.e. and take on only a finite set of values. B_0 is not sequentially complete. We define $x(s) \equiv x_t(s) = 1$ for $t \geqq s$, and to vanish elsewhere. $x(s)$ is integrable on σ to $(|\sigma \cdot I_0{}^t|)$. However if $\phi(s) = s$, then $\int_0^1 \phi(s)x(s)ds = (t^2/2)$ which is not in B_0.

10.7. *Example showing the need of completeness with respect to D in a space which is sequentially complete* (*Theorem* 4.1). Let l_0 be the linear space of functions x_t on $T = (0, 1)$ such that $\sum_t |x_t| < \infty$ and let $\bar{l}_0$ be the set of all

bounded measurable functions $\bar{x}_t$ on $T=(0, 1)$. Then $\bar{x}(x)\equiv\sum\bar{x}_t\cdot x_t$ is a linear continuous operation on l_0. We will define the topology on l_0 by all finite sets of elements from $\bar{l}_0$. Any denumerable sequence of elements of l_0 differs on T at only a denumerable set of points. l_0 is sequentially complete since this is equivalent to the weak completeness of l_1 [6, p. 143]. Finally set $x_t(s)=\delta_{st}$ on $S=(0, 1)$ to l_0. As $\int_\sigma\bar{x}(x_t(s))ds=\int_\sigma\bar{x}_t dt$ exists, it follows that for every $U\in\mathcal{U}$ (finite set of $\bar{x}$'s) there will be a $J_\Delta(x, S)$ which is unconditionally summable with respect to U. On the other hand there exists no $x^0\in l_0$ such that $\sum_t\bar{x}_t\cdot\bar{x}_t{}^0=\int_0^1\bar{x}_t dt$ for all $\bar{x}\in\bar{l}_0$. $x(s)$ is therefore not $\mathcal{U}$-integrable.

For convenience let M be the class of functions $x(s)$ on $S=(0, 1)$ to B such that (1) $\bar{x}[x(s)]$ as measurable for all $\bar{x}\in\bar{B}$, and (2) $\int_\sigma\bar{x}[x(s)]ds$ are uniformly absolutely continuous with respect to elements of $\bar{B}$. If $x(s)$ is contained in a separable subspace of B, then it is Pettis integrable [4, p. 293, Theorem 5.3]. It is Dunford integrable in any case.

10.8. *Example of a member of M which is not Pettis integrable.* Sierpiński [21, pp. 9–10] has shown that equivalent to the hypothesis of the continuum is the existence of a set θ in the unit square I such that on every parallel to the s axis θ contains at most a denumerable set of points in I while on every parallel to the t axis $I-\theta$ contains at most a denumerable set of points in I. We define $x(s)$ on $S=(0, 1)$ to B so that $x(s)\equiv x_t(s)$ is the characteristic function of θ. In the notation of Example 10.2,

$$\bar{x}[x(s)]=\int_T x_t(s)d\beta=\int_T x_t(s)d\beta_1+\int_T x_t(s)d\beta_2.$$

Since for any $t\in(0, 1)$, $x_t(s)=1$ for only a denumerable set δ_t in S, it follows that $\int_T x_t(s)d\beta_1\neq 0$ on at most the set $\delta_0=\sum_{t\in R}\delta_t$ which is denumerable. Let 1_t be the element of B which has the value one for all $t\in T$. As $(1_t-x_t(s))$ has only a denumerable set of nonzero values for each $s\in S$, we have $\int_T(1_t-x_t(s))d\beta_2=0$. Hence $\bar{x}[x(s)]=\int_T 1_t d\beta_2$ for all $s\in(S-\delta_0)$ so that $\bar{x}[x(s)]$ is measurable. The integrals $\int_\sigma\bar{x}[x(s)]ds$ are obviously uniformly absolutely continuous with respect to elements of $\bar{B}$. Thus $x(s)\in M$.

Finally, let us suppose that there exists a $y\in B$ such that $\bar{x}(y)=\int_s\bar{x}[x(s)]ds$ for every $\bar{x}\in B$. For every $t_0\in T$ there exists an $\bar{x}_{t_0}\in\bar{B}$ defined by $\bar{x}_{t_0}(x_t)=x_{t_0}$. This set is total on B.

$$y_{t_0}=\bar{x}_{t_0}(y)=\int_S\bar{x}_{t_0}[x_t(s)]ds=\int_S x_{t_0}(s)ds=0.$$

Now Banach [6, p. 30] has demonstrated the existence of $\bar{x}_0\in\bar{B}$ for which $\bar{x}_0(x)=\mathrm{L}\int_s x_t dt$ whenever x_t is Lebesgue integrable. Our supposition then leads to

$$0 = \bar{x}_0(y) = \int_S \bar{x}_0[x_i(s)]ds = \int_S 1ds = 1.$$

Therefore $x(s)$ is not Pettis integrable.

It is to be noticed that $F(\bar{x}) = \int_\epsilon \bar{x}[x(s)]ds$ is weakly continuous on $\bar{B}$. Since there exists no $y \in B$ such that $F(\bar{x}) = \bar{x}(y)$ for all $\bar{x} \in \bar{B}$, the Banach theorem [6, p. 131, Theorem 8] to this effect for separable Banach spaces is not in general true for nonseparable Banach spaces.

10.9. *Example of an additive, absolutely continuous function $x(\sigma)$ on measurable sets of* $(0, 1)$ *to Hilbert space H which in the weak neighborhood topology is pseudo-differentiable but a.e. not differentiable* (§8). Let $x_{i,j}$ be a doubly infinite set of orthonormal vectors in H. We define $y_i(s) = 2^i x_{i,j}$ on $(j/2^i, 2^{-2i}+j/2^i)$ $j = 0, 1, \cdots, (2^i-1)$, and elsewhere set $y_i(s) = 0$. As Birkhoff [1, p. 375, Example 2] has observed, $y(s) = \sum_{n=1}^{\infty} y_n(s)$ exists a.e. and is Birkhoff integrable, Pettis has shown [4] that $x(\sigma) = \int_\sigma y(s)ds$ is additive, absolutely continuous and in the weak neighborhood topology pseudo-differentiable. We will now demonstrate that $\|x(I)/|I|\|$ is unbounded a.e. As in Theorem 8.3, this will imply that a.e. the weak derivative does not exist. Let $s = {}_0a_1a_2 \cdots a_n \cdots$ be the dyadic representation of $s \in (0, 1)$. Hardy and Littlewood [22, p. 189] have proved that any finite sequence of digits will recur an infinite number of times on a set of measure one. In particular, n consecutive zeros will occur infinitely often on σ_n of unit measure. Then $\sigma_0 = \prod \sigma_n$ is of measure one. For each point of σ_0 all numbers of consecutive zeros occur infinitely often. Suppose p is the number of consecutive zeros not exceeding $n/2$ which immediately proceed the $(n+1)$st digit of ${}_0a_1a_2 \cdots a_n \cdots = s$. If I_n is the interval ${}_0a_1a_2 \cdots a_{n-p} \leqq s \leqq {}_0a_1a_2 \cdots a_{n-p} + 1/2^n$, then

$$x(I_n) = \int_{I_n} \left[\sum_{n-p}^{n} y_i(s) \right] ds + \int_{I_n} \left[y(s) - \sum_{n-p}^{n} y_i(s) \right] ds$$

$$= \sum_{n-p}^{n} x_{i,j_i}/2^i + x'(I_n)$$

where $x'(I_n)$ is orthogonal to (x_{i,j_i}) $(i = n-p, \cdots, n)$. In this case $\|x(I_n)/|I_n|\| \geqq 2^p$. Since for each point of σ_0, $\lim_n p_n = \infty$, our assertion is proved.

References

1. G. Birkhoff, *Integration of functions with values in a Banach space*, these Transactions, vol. 38 (1935), pp. 357–378.

2. N. Dunford, *Uniformity in linear spaces*, these Transactions, vol. 44 (1938), pp. 305–356.

3. I. Gelfand, *Sur un lemme de la théorie des espaces linéaires*, Communications de l'Institut des Sciences Mathématiques et Mécaniques de l'Université de Kharkoff et la Société Mathématique de Kharkoff, vol. 13, pp. 35–40.

4. B. J. Pettis, *On integration in vector spaces*, these Transactions, vol. 44 (1938), pp. 277–304.

5. A. Kolmogoroff, *Zur Normierbarkeit eines allgemeinen topologischen linearen Raumes*, Studia Mathematica, vol. 5 (1934), pp. 29–33.

6. S. Banach, *Théorie des Opérations Linéaires*, Monografje Matematyczne, Warsaw, 1932.

7. J. von Neumann, *On complete topological spaces*, these Transactions, vol. 37 (1935), pp. 1–20.

8. J. V. Wehausen, *Transformations in linear topological spaces*, Duke Mathematical Journal, vol. 4 (1938), pp. 157–169.

9. E. H. Moore, *General Analysis*, Memoirs of the American Philosophical Society, vol. 1, part 2, 1939.

10. E. H. Moore and H. L. Smith, *A general theory of limits*, American Journal of Mathematics, vol. 44 (1922), p. 103.

11. Fréchet, *Sur l'intégrale d'une fonctionelle étendue à un ensemble abstrait*, Bulletin de la Société Mathématique de France, vol. 43 (1915), pp. 248–265.

12. S. Saks, *Addition to the note on some functionals*, these Transactions, vol. 35 (1933), p. 967.

13. S. Bochner, *Integration von Funktionen deren Werte die Elemente eines Vektorraumes sind*, Fundamenta Mathematicae, vol. 20 (1933), pp. 262–276.

14. J. A. Clarkson, *Uniformly convex spaces*, these Transactions, vol. 40 (1936), pp. 396–414.

15. B. J. Pettis, *A note on regular Banach spaces*, Bulletin of the American Mathematical Society, vol. 44 (1938), pp. 420–428.

16. N. Dunford and A. P. Morse, *Remarks on the preceding paper of James A. Clarkson*, these Transactions, vol. 40 (1936), pp. 415–420.

17. C. Carathéodory, *Vorlesungen über reelle Funktionen*, Leipzig, 1918.

18. G. Birkhoff, *A note on topological groups*, Compositio Mathematica, vol. 3 (1936), pp. 427–430.

19. T. H. Hildebrandt, *Linear operations on functions of bounded variation*, Bulletin of the American Mathematical Society, vol. 44 (1938), p. 75.

20. K. Knopp, *Mengentheoretische Behandlung einiger Probleme der diophantischen Approximationen und der transfiniten Wahrscheinlichkeiten*, Mathematische Annalen, vol. 95 (1925), pp. 409–420.

21. W. Sierpiński, *Hypothèse du Continu*, Monografje Matematyczne, Warsaw, 1934.

22. G. H. Hardy and J. E. Littlewood, *Some problems of diophantine approximation*, Acta Mathematica, vol. 37 (1914), pp. 155–190.

23. B. J. Pettis, *Differentiation in Banach spaces*, Duke Mathematical Journal, vol. 5 (1939), pp. 254–269.

UNIVERSITY OF MICHIGAN,
ANN ARBOR, MICH.

HARMONIC SURFACES IN RIEMANN METRIC*

BY

S. BOCHNER

If x^i $(i=1, \cdots, n)$ is a vector representing a map from a domain Q in euclidean (u, v)-space into the euclidean x^i-space, the Euler equations of the problem

$$\iint (x^i_u x^i_u + x^i_v x^i_v)dudv = \min, \qquad \text{summation over } i, \tag{1}$$

have the simple form

$$\Delta x^i \equiv x^i_{uu} + x^i_{vv} = 0. \tag{2}$$

The boundary value problem reduces trivially to the classical Dirichlet problem in its simplest form. If the x^i-space is a Riemannian space S with a positive definite metric, the generalization of (1) is

$$\iint g_{pq}(x)(x^p_u x^q_u + x^p_v x^q_v)dudv = \min,$$

and the Euler equations are

$$\frac{\partial g_{pq}}{\partial x^i}(x^p_u x^q_u + x^p_v x^q_v) - 2\frac{\partial}{\partial u}(g_{iq}x^q_u) - 2\frac{\partial}{\partial v}(g_{iq}x^q_v) = 0.$$

Putting

$$\frac{\partial g_{pq}}{\partial x^i} = {}_{\delta pa}\Gamma^a_{iq} + g_{aq}\Gamma^a_{pi}, \tag{3}$$

we obtain the equivalent equations

$$\Delta x^i + \Gamma^i_{pq}(x^p_u x^q_u + x^p_v x^q_v) = 0. \tag{4}$$

Any solution of (4) will be called a *harmonic surface.*

Equations (4) are nonlinear in the first derivatives of x^i and the boundary value problem is now decidedly more difficult than in the classical case (2). There is no direct generalization of Poisson's integral and the results obtained directly by Picard's approximation method are geometrically very restricted.†

* Presented to the Society, October 28, 1939; received by the editors June 2, 1939.

† G. Giraud, *Sur le problème de Dirichlet généralisé*, II, Annales de l'École Normale Supérieure, vol. 46 (1929), pp. 131–245.

However S. Bernstein's elaborate method of majorization* can be shown to be effective in our case, and in the present note we shall offer some first results in this direction. In §1 we shall state some elementary properties of harmonic surfaces and in §2 a uniqueness theorem relative to the boundary value problem; and in §3 we shall prove a theorem on compactness of families of harmonic surfaces.

1. **Subharmonic character of harmonic surfaces.** Equations (4) are an extension, from one to two parameters, of the equations for geodesics:

$$\frac{d^2x^i}{ds^2} + \Gamma^i_{pq}\frac{dx^p}{ds}\frac{dx^q}{ds} = 0. \tag{5}$$

As in the latter case, equations (4) can be easily seen to be formally invariant under transformations of coordinates in the x^i-space, the expressions on the left side of (4) being the components of a vector relative to such transformations. Other quantities which are formally invariant are the coefficients

$$E = g_{pq}x^p_u x^q_u, \qquad F = g_{pq}x^p_u x^q_v, \qquad G = g_{pq}x^p_v x^q_v \tag{6}$$

of the fundamental form $Edu^2+2Fdudv+Gdv^2$. These coefficients are scalars.

As in the euclidean case, the expression

$$E - G - 2F(-1)^{1/2} = g_{pq}(x)(x^p_u - (-1)^{1/2}x^p_v)(x^q_u - (-1)^{1/2}x^q_v)$$

is an analytic function of the complex variable $u+iv$. In other words we have the Cauchy-Riemann equations

$$E_u - G_u = -2F_v, \qquad E_v - G_v = 2F_u, \tag{7}$$

$$\Delta E = \Delta G, \qquad \Delta F = 0. \tag{8}$$

Relations (7) are formal identities. They hold for expressions (6), subject to relations (3) and (4). The differentiations involved can be reduced to a minimum by the use of geodesic coordinates. The first derivatives of E, F, G with respect to u and v involve the coefficients g_{pq}, Γ^i_{pq} but no derivatives of Γ^i_{pq}. In order to prove (7) for a fixed value (u_0, v_0) we may select geodesic coordinates originating at the point $x^i(u_0, v_0)$, for which, at the origin,

$$\Gamma^i_{pq} = 0, \qquad g_{pq} = \delta_{pq}. \tag{9}$$

Thus the proof of (7) consists of proving the same trivial identities as in the euclidean case.

* S. Bernstein, *Sur la généralisation du problème de Dirichlet*, Mathematische Annalen, vol. 69 (1910), pp. 82–136, §1.

If S is a space whose Riemannian curvature is nonpositive, then relations (8) can be supplemented by the further relation

$$\Delta E \geqq 0, \qquad \Delta G \geqq 0. \tag{10}$$

We shall derive a more comprehensive result relating to harmonic "surfaces" in several dimensions. If we replace the two parameters u, v by parameters u^α, $\alpha = 1, \cdots, m$, and put

$$\Delta x = \frac{\partial^2 x}{\partial u^\alpha \partial u^\alpha} \equiv \sum_{\alpha=1}^{m} \frac{\partial^2 x}{\partial u^\alpha \partial u^\alpha}, \qquad E_{\beta\gamma} = g_{pq} \frac{\partial x^p}{\partial u^\beta} \frac{\partial x^q}{\partial u^\gamma},$$

then the equations

$$\Delta x^i + \Gamma^i_{pq} \frac{\partial x^p}{\partial u^\alpha} \frac{\partial x^q}{\partial u^\alpha} = 0$$

are the Euler equations of the problem

$$\int\int E_{\alpha\alpha}\, du_1 du_2 \cdots du_m \equiv \min.$$

If for given values u^α, we select a coordinate system which is geodesic at the point $x^i(u^\alpha)$, then we obtain

$$\frac{\partial E_{\beta\gamma}}{\partial u^\alpha} = g_{pq}\left(\frac{\partial^2 x^p}{\partial u^\beta \partial u^\alpha} \frac{\partial x^q}{\partial u^\gamma} + \frac{\partial x^p}{\partial u^\beta} \frac{\partial^2 x^q}{\partial u^\gamma \partial u^\alpha}\right) + g_{pa}\Gamma^a_{qr}\left(\frac{\partial x^p}{\partial u^\beta} \frac{\partial x^q}{\partial u^\gamma} + \frac{\partial x^p}{\partial u^\gamma} \frac{\partial x^q}{\partial u^\beta}\right) \frac{\partial x^r}{\partial u^\alpha}$$

and hence, putting $\Gamma^i_{pq} = 0$,

$$\Delta E_{\beta\gamma} = g_{pq}\left(\Delta \frac{\partial x^p}{\partial u^\beta} \cdot \frac{\partial x^q}{\partial u^\gamma} + \frac{\partial x^p}{\partial u^\beta} \cdot \Delta \frac{\partial x^q}{\partial u^\gamma}\right) + 2g_{pq} \frac{\partial^2 x^p}{\partial u^\beta \partial u^\alpha} \frac{\partial^2 x^q}{\partial u^\gamma \partial u^\alpha}$$
$$+ g_{pa} \frac{\partial \Gamma^a_{qr}}{\partial x^s} \frac{\partial x^s}{\partial u^\alpha}\left(\frac{\partial x^p}{\partial u^\beta} \frac{\partial x^q}{\partial u^\gamma} + \frac{\partial x^p}{\partial u^\gamma} \frac{\partial x^q}{\partial u^\beta}\right) \frac{\partial x^r}{\partial u^\alpha}.$$

Now, using the formula

$$\frac{\partial}{\partial u^\beta} \Delta x^p = -\frac{\partial \Gamma^p_{ab}}{\partial x^q} \frac{\partial x^q}{\partial u^\beta} \frac{\partial x^a}{\partial u^\alpha} \frac{\partial x^b}{\partial u^\alpha}$$

and a similar formula for $\partial \Delta x^q / \partial u^\gamma$ and putting $R_{pqrs} = g_{pa}(\partial \Gamma^a_{rs}/\partial x^q - \partial \Gamma^a_{rq}/\partial x^s)$, we finally obtain

$$\Delta E_{\beta\gamma} = 2g_{pq} \frac{\partial^2 x^p}{\partial u^\beta \partial u^\alpha} \frac{\partial^2 x^q}{\partial u^\gamma \partial u^\alpha} - R_{pqrs}\left(\frac{\partial x^p}{\partial u^\beta} \frac{\partial x^q}{\partial u^\gamma} + \frac{\partial x^q}{\partial u^\beta} \frac{\partial x^p}{\partial u^\gamma}\right) \frac{\partial x^r}{\partial u^\alpha} \frac{\partial x^s}{\partial u^\alpha}.$$

For any quantities λ^α, putting $\mu_\alpha{}^p = \lambda^\beta(\partial^2 x^p/\partial u^\beta \partial u^\alpha)$, $\nu^p = \lambda^\beta(\partial x^p/\partial u^\beta)$, we hence obtain

$$\Delta E_{\beta\gamma}\lambda^\beta\lambda^\gamma = 2g_{pq}\mu_\alpha{}^p\mu_\alpha{}^q - 2R_{pqrs}\nu^p{}^q\nu \frac{\partial x^r}{\partial u^\alpha}\frac{\partial x^s}{\partial u^\alpha}.$$

Thus, for a space S of nonpositive curvature, $\Delta E_{\beta\gamma}\lambda^\beta\lambda^\gamma \geqq 0$, and this implies relations (10).

Finally we shall indicate a proof for the assertion that in a space S of nonpositive curvature, the curvature $K(u, v)$ of any harmonic surface is also nonpositive. By a general theorem of Ricci,* K is the sum of two quantities K_1 and K_2; K_1 is the so-called relative curvature of the surface and K_2 is the curvature of S at the point $x^i(u, v)$. Since $K_2 \leqq 0$ by assumption, we have only to show that $K_1 \leqq 0$. Now, if (9) holds, the quantity K_1 can be computed by a formula which is formally the same as in the euclidean case, namely

$$(EG - F^2)K_1 = \begin{vmatrix} x^i_{uu}, & x^i_u, & x^i_v \end{vmatrix} \begin{vmatrix} x^i_{vv}, & x^i_u, & x^i_v \end{vmatrix} - \begin{vmatrix} x^i_{uv}, & x^i_u, & x^i_v \end{vmatrix} \begin{vmatrix} x^i_{uv}, & x^i_u, & x^i_v \end{vmatrix}.$$

Each term on the right side is the (numerical) product of two matrices of 3 rows and n columns each. Now, (9) implies $x^i_{vv} = -x^i_{uu}$ and therefore $K_1 \leqq 0$.

2. **The distance function.** If two solutions of (2) over the same region Q have the same limit values on the boundary B of Q, then they are identical throughout Q. We must not expect a similar theorem to hold without restriction for solutions of (4) since it does not even hold for solutions of the simpler equations (5). However, the restrictions required for equations (4) are exactly the same as those needed for equations (5). It is sufficient to assume that the solutions lie in a region U of the space S in which there are no conjugate points. More precisely we assume that any two points x^i and y^i of U can be joined in S by a geodesic arc whose closure contains no conjugate points and whose geodesic length is the shortest geodesic distance between the points x^i and y^i. Denoting the latter distance by $H(x^i, y^i)$, if the space S belongs to class C_4, then $H(x^i, y^i)$ belongs to class C_2 on the domain $U \times U$ minus the diagonal manifold $x^i = y^i$.

For an arbitrary function $W = W(x^i, y^i)$ and arbitrary numbers ξ^i, η^i we shall consider the expression

$$(W_{x^p x^q} - \Gamma^i_{pq}(x)W_{x^i})\xi^p{}^q\xi + 2W_{x^p y^q}\xi^p{}^q\eta + (W_{y^p y^q} - \Gamma^i_{pq}(y)W_{y^i})\eta^p{}^q\eta$$

(the subscripts of W indicating partial derivations) and we shall denote this expression by

$$P(W) \equiv P(W; \xi, \eta).$$

* L. P. Eisenhart, *Riemannian Geometry*, 1926, p. 174.

We shall say that *a function $W(x^i, y^i)$ has positive character, if $P(W; \xi, \eta) \geqq 0$ identically in $x^i, y^i; \xi^i, \eta^i$ (assuming $x^i \neq y^i$, if necessary).*

We shall now prove

LEMMA 1. *The distance function $H(x^i, y^i)$ has positive character.*

We take a geodesic arc issuing from x^i, whose direction cosines at the point of issuance have the given values ξ^i, and on it a point of geodesic distance ϵ. Denoting the coordinates of the latter point by $x^i + \xi^i(\epsilon)$ we will have by familiar formulas from the theory of normal coordinates,*

$$\xi^i(\epsilon) = \epsilon\xi^i - \frac{\epsilon^2}{2!}\Gamma^i_{pq}(x)\xi^p\xi^q + o(\epsilon^2) \tag{11}$$

and

$$H(x^i, x^i + \xi^i(\epsilon)) = \epsilon\left\{g_{pq}(x)\xi^p\xi^q\right\}^{1/2}.$$

In particular

$$\frac{d}{d\epsilon}H(x^i, x^i + \xi^i(\epsilon)) = 0. \tag{12}$$

Similarly we put

$$\eta^i(\epsilon) = \epsilon\eta^i - \frac{\epsilon^2}{2!}\Gamma^i_{pq}(y)\eta^p\eta^q + o(\epsilon^2) \tag{13}$$

and obtain

$$\frac{d}{d\epsilon}H(y^i, y^i + \eta^i(\epsilon)) = 0. \tag{14}$$

Since there are no conjugate points on the geodesic arc joining x^i and y^i, the second derivative with respect to ϵ of

$$H(x^i, x^i + \xi^i(\epsilon)) + H(x^i + \xi^i(\epsilon), y^i + \eta^i(\epsilon)) + H(y^i, y^i + \eta^i(\epsilon))$$

is greater than or equal to zero for $\epsilon = 0$. On account of (12) and (14), we therefore have

$$\lim_{\epsilon \to 0}\frac{d^2}{d\epsilon^2}H(x^i + \xi^i(\epsilon), y^i + \eta^i(\epsilon)) \geqq 0.$$

Substituting (11) and (13) we hence obtain $P(H; \xi, \eta) \geqq 0$, which is the assertion of our lemma.

LEMMA 2. *If $\chi(t)$ is differentiable twice and $\chi'(t) \geqq 0$ and $\chi''(t) \geqq 0$, and if $W(x^i, y^i)$ has positive character, then $\chi(W(x^i, y^i))$ has positive character too.*

* T. Y. Thomas, *Differential Invariants of Generalized Spaces*, 1934, p. 7.

In fact,

$$P(\chi(W); \xi, \eta) = \chi'(W) \cdot P(\chi(W); \xi, \eta) + \chi''(W) \cdot (W_{x^p}\xi^p + W_{y^p}\eta^p)^2.$$

In particular, we see that the function $(H(x^i, y^i))^2$ has positive character, and this function has the advantage of belonging to class C_2 throughout $U \times U$, the manifold $x^i = y^i$ not being excepted.

LEMMA 3. *If $W(x^i, y^i)$ has positive character and $x^i(u, v)$, $y^i(u, v)$ are two harmonic surfaces, then the function $f(u, v) = W(x^i(u, v), y^i(u, v))$ is subharmonic, that is $\Delta f \geqq 0$.*

In fact, carrying out the differentiations and replacing Δx^i and Δy^i by $-\Gamma^i_{pq}(x)(x_u{}^p x_u{}^q + x_v{}^p x_v{}^q)$ and $-\Gamma^i_{pq}(y)(y_u{}^p y_u{}^q + y_v{}^p y_v{}^q)$ respectively, we obtain

$$\Delta f = P(W; x_u, y_u) + P(W; x_v, y_v),$$

and therefore $\Delta f \geqq 0$.

Since a subharmonic function attains its maximum on the boundary and since the distance function $H(x^i, y^i)$ vanishes only for $x^i = y^i$, we may conclude the following theorem.

THEOREM 1. *If U is a domain of S on which the distance function $H(x^i, y^i)$ has positive character, then for any two harmonic surfaces $x^i(u, v)$, $y^i(u, v)$, the function $H(x^i(u, v), y^i(u, v))$ attains its maximum on the boundary B of Q. If $x^i(u, v) = y^i(u, v)$ on B, then this equality holds throughout $Q + B$.*

If $K(a^i, \sigma)$ is a geodesic sphere $H(a^i, x^i) < \sigma$ in U, then any harmonic surface in U whose boundary lies in $K(a^i, \sigma)$ lies there in its entirety.

The last part of the theorem is a consequence of the first if we put $y^i(u, v) \equiv a^i$.

3. **The Dirichlet integrand on the boundary.** Using polar coordinates $u = \rho \cos \theta$, $v = \rho \sin \theta$, we assume now that Q is the unit circle $\rho < 1$, $0 \leqq \theta < 2\pi$, and we shall denote the same function in Q by $x(u, v)$ and $x(\rho, \theta)$ interchangeably. If a periodic function $\phi(\theta)$ is continuous, and has m continuous derivatives, $m = 0, 1, \cdots$, we shall write

$$\|\phi(\theta)\|_m = \max_{0 \leqq \theta < 2\pi} \left| d^m \phi(\theta)/d\theta^m \right|.$$

In case of a vector $\phi^i(\theta)$ we shall write

$$\|\phi(\theta)\|_m = \sum_{i=1}^{n} \|\phi^i(\theta)\|_m.$$

This symbolism will refer exclusively to the variable denoted by θ.

The theorem to follow will apply to a sufficiently small neighborhood U

of a fixed point on S in which the distance function has positive character. Making U sufficiently small and choosing the coordinate system appropriately, we may first assume that U is the entire cell $-\infty < x^i < \infty$, $i = 1, \cdots, n$, the fixed point of U being the origin $x^i = 0$. U contains some geodesic sphere $K(0, \sigma')$ whose radius will be also denoted by $3\sigma^0$. Obviously, for $0 < \sigma < 3\sigma^0$ there exist two functions $r'(\sigma)$ and $r''(\sigma)$, $0 < r' < r''$, such that the geodesic sphere $K(0, \sigma)$ contains the euclidean sphere $x^i x^i \leqq r'^2$ and is in its turn contained in the euclidean sphere $x^i x^i \leqq r''^2$. We further observe that for $x^i(u, v)$ in $K(0, 2\sigma^0)$, the quantity

$$E + G = g_{pq}(x)(x_u^p x_u^q + x_v^p x_v^q) \tag{15}$$

is majorized above and below by

$$x_u^i x_u^i + x_v^i x_v^i. \tag{16}$$

Now, for $a^i \subset K(0, 2\sigma^0)$ and $x^i \subset K(0, \sigma^0)$ we put $\xi^i = x^i - a^i$, and we consider the function $W(a^i, \xi^i) = (H(a^i, x^i))^2/2$, and the quantities

$$\eta^i = \eta^i(a^p, \xi^q) = W_{\xi^i}. \tag{17}$$

Since, for $\xi^i = 0$, $\partial \eta^p / d\xi^q = g_{pq}(a)$, we may consider the inverse functions

$$\xi^i = \xi^i(a^p, \eta^q). \tag{18}$$

They are defined, and they are inverse to the functions (17), for

$$a^i \subset K(0, 2\sigma^0), \qquad \eta^i \eta^i \leqq r_0^2 \qquad (> 0), \tag{19}$$

the number r_0 being independent of a^i.

We now consider a solution $x^i(u, v)$ of (4) in Q which lies in $K(0, \sigma^0)$ and whose partial derivatives are uniformly continuous in Q. The functions $x^i(u, v)$ and their partial derivatives of first and second order have continuous boundary values on B. The function $f(u, v) = f(\rho, \theta) = W(a^i, x^i(u, v) - a^i)$ has partial derivatives of second order which are uniformly continuous in Q, and satisfies

$$\|f(1, \theta)\|_m \leqq M \|x(1, \theta)\|_m, \qquad m = 0, 1, 2,$$

the constant M being independent of $x^i(u, v)$ but not necessarily the same in all relations in which it will occur. The important property of $f(u, v)$ is that it is subharmonic, $\Delta f \geqq 0$ (see Theorem 1). We now put

$$f(\rho, \theta) = g(\rho, \theta) + h(\rho, \theta)$$

where $g(\rho, \theta)$ is the harmonic function with the boundary values $f(1, \theta)$ and

$h(\rho, \theta)$ is a subharmonic function with boundary values 0. For $g(\rho, \theta)$ we have*

$$\left|\frac{\partial g(\rho, \theta)}{\partial \rho}\right| + \left|\frac{\partial g(\rho, \theta)}{\partial \theta}\right| \leqq M\|x(1, \theta)\|_2.$$

For the function $h(\rho, \theta)$ we immediately obtain the relation

$$\lim_{\rho \to 1} \left|\frac{\partial h(\rho, \theta)}{\partial \theta}\right| \equiv \left|\frac{\partial h(1, \theta)}{\partial \theta}\right| = 0,$$

which is an immediate consequence of the assumption that $h(\rho, \theta)$ shall vanish on the boundary. For the radial derivative we obtain the one-sided relation

$$\lim_{\rho \to 1} \frac{\partial h(\rho, \theta)}{\partial \rho} \equiv \frac{\partial h(1, \theta)}{\partial \rho} \geqq 0$$

with no absolute values to start with. It follows from the fact that $h(\rho, \theta)$ is zero on the boundary and, being subharmonic, is nonpositive in Q. Altogether we have the estimates

$$\left|\frac{\partial f(1, \theta)}{\partial \theta}\right| \leqq M\|x(1, \theta)\|_2, \quad -\frac{\partial f(1, \theta)}{\partial \rho} \leqq M\|x(1, \theta)\|_2,$$

and this can also be written in the form

$$\max\left|\eta^i(1, \theta)\frac{\partial x^i(1, \theta)}{\partial \theta}\right| - \eta^i(1, \theta)\frac{\partial x^i(1, \theta)}{\partial \rho} \leqq M\|x(1, \theta)\|_2.$$

The constant M is independent of θ and a^i, and all combinations (19) are admissible. Hence we obtain the estimate

$$\left|\frac{\partial x^i(1, \theta)}{\partial \theta}\right| + \left|\frac{\partial x^i(1, \theta)}{\partial \rho}\right| \leqq M\|x(1, \theta)\|_2,$$

the second quantity on the left side being also an absolute value. Since

$$\left(\frac{\partial x^i}{\partial u}\right)^2 + \left(\frac{\partial x^i}{\partial v}\right)^2 = \left(\frac{\partial x^i}{\partial \rho}\right)^2 + \frac{1}{\rho^2}\left(\frac{\partial x^i}{\partial \theta}\right)^2,$$

we finally obtain the following theorem.

THEOREM 2. *If S belongs to class C_4, if U is a sufficiently small coordinate neighborhood of the origin, and if $x^i(u, v) = x^i(\rho, \theta)$ is a harmonic surface in U*

* H. Whitney, *Analytic coordinate systems in a manifold*, Annals of Mathematics, (2), vol. 38 (1937), p. 809, Lemma 4.

whose partial derivatives of the second order are uniformly continuous in the unit circle Q, then the inequality

$$\{E(1,\theta)+G(1,\theta)\}^{1/2} \leqq M\sum_{i=1}^{n} \max \left|\frac{\partial^2 x^i(1,\theta)}{\partial\theta^2}\right|$$

holds for a constant M which is independent of the given surface.

If S has nonpositive curvature, then the sharper inequality

$$\{E(u,v)+G(u,v)\}^{1/2} \leqq M\sum_{i=1}^{n} \max \left|\frac{\partial^2 x^i(1,\theta)}{\partial\theta^2}\right| \tag{20}$$

holds for the same constant M.

The second half of the theorem follows from the first since for spaces of nonpositive curvature, $E+G$ attains its maximum on the boundary of Q. But the inequality (20) probably holds for spaces S of any curvature.

From Theorem 2 we could easily deduce the existence theorem* that any vector $\phi^i(\theta)$ in U which has two continuous derivatives is the boundary of a harmonic surface in U.

---

* J. Léray et J. Schauder, *Topologie et équations fonctionelles*, Annales de l'École Normale Supérieure, vol. 51 (1934), pp. 45–78.

PRINCETON UNIVERSITY,
PRINCETON, N. J.

A CONVERGENCE THEOREM FOR CONTINUED FRACTIONS*

BY

W. T. SCOTT AND H. S. WALL

1. **Introduction.** This paper is concerned with the question of convergence of continued fractions of the form

$$\frac{1}{1}+\frac{a_2}{1}+\frac{a_3}{1}+\frac{a_4}{1}+\cdots \tag{1.1}$$

in which the a_n are complex numbers. Our investigation began with a comparison of the infinite series

$$\frac{A_1}{B_1}+\left(\frac{A_2}{B_2}-\frac{A_1}{B_1}\right)+\left(\frac{A_3}{B_3}-\frac{A_2}{B_2}\right)+\cdots, \tag{1.2}$$

equivalent to the sequence of approximants of the continued fraction, with a convergent positive term series $\sum p_n$. It was found that if $r_n=p_{n+1}/p_n$, and

$$r_n\left|1+a_n+a_{n+1}\right|\geqq r_nr_{n-2}\left|a_n\right|+\left|a_{n+1}\right|, \tag{1.3}$$

$n=1, 2, 3, \cdots$, $r_0=r_{-1}=a_1=0$, then the series $\sum p_n$ is a majorant for (1.2), so that the continued fraction converges.

Later, in proving that the Pringsheim convergence theorem can be obtained from the above by specializing the series $\sum p_n$, it was discovered that the inequalities (1.3) have a second interpretation. In fact, if $a_n\neq 0$, and

$$y_0+\frac{x_1}{y_1}+\frac{x_2}{y_2}+\frac{x_3}{y_3}+\cdots \tag{1.4}$$

denotes the even (odd) "contraction" of (1.1), then the inequalities (1.3) for odd (even) values of n may be written in the form

$$\left|c_1y_1\right|\geqq 1, \qquad \left|c_ny_n\right|\geqq\left|c_{n-1}c_ny_n\right|+1, \tag{1.5}$$

$n=2, 3, 4, \cdots$, where $c_n\neq 0$. But these are precisely the *Pringsheim relations*† for the continued fraction (1.4). Hence if the inequalities (1.3) hold for some $r_n>0$ (not necessarily related to a convergent series $\sum p_n$), the

* Presented to the Society, September 8, 1939; received by the editors May 26, 1939, and, in revised form, October 27, 1939.

† O. Perron, *Die Lehre von den Kettenbrüchen* (referred to hereafter as "Perron"), 2d edition, Leipzig and Berlin, 1929, pp. 257–258, formulas (5).

sequences of even and odd approximants of (1.1) have limits, finite or infinite, and the limit of the even (odd) approximants is finite if inequality holds in (1.3) for an odd (even) index. *Thus the relations* (1.3) *have two interpretations, and therefore are extremely important instruments in the convergence theory.*

One of the most important results obtained in the paper is the "parabola theorem," namely, that if the elements a_n of (1.1) lie within or upon the parabola $|z|-\Re(z)=1/2$, then the continued fraction converges if and only if some a_n is 0, or the series $\sum|b_n|$ diverges, where $b_1=1$, $a_n=1/b_{n-1}b_n$, $n=2, 3, 4, \cdots$. In the first announcement of the theorem,* the a_n were restricted to lie in a closed bounded region entirely within the parabola, a result which was obtained using $r_n=r<1$ in (1.3). Using the *second* interpretation of the inequalities, we were able later to remove the restriction requiring the a_n to lie in the interior of the parabola, and to replace the boundedness condition by the condition on the series $\sum|b_n|$.

The parabola cannot be replaced by a "better" curve symmetric with respect to the real axis. *We thus characterize completely all regions in the complex plane, symmetric with respect to the real axis, in which the elements a_n may vary independently while the continued fraction remains convergent.* We have therefore, in a certain sense, extended as far as possible the theorem given by Worpitzky† three-quarters of a century ago, namely, that (1.1) converges if $|a_n|\leqq 1/4$, $n=2, 3, 4, \cdots$.

Other results obtained include: a notable improvement in the Pringsheim theorem on uniform convergence (Theorem B), the "triangle theorem" (Theorem G), some results on the *problem of Szász* (§6), and the theorem that if (1.3) holds and the a_n are bounded, then (1.1) converges (Theorem E).

The methods used throughout the paper are elementary in character, and are based strictly upon the continued fraction algorithm.

2. **The first interpretation of the fundamental inequalities.** The basic result of this paper is embodied in the theorem which follows.

THEOREM A. *Let* (1.1) *be a continued fraction with complex elements a_n, and nth approximant A_n/B_n. Suppose that for some nonnegative numbers r_n the inequalities* (1.3) *hold. Then $B_n\neq 0$, $n=1, 2, 3, \cdots$, and the series $1+\sum r_1r_2\cdots r_n$ is a majorant for the series* (1.2). *The continued fraction converges if some a_n is 0, or if the series $1+\sum r_1r_2\cdots r_n$ converges. The sum of the latter series is an upper bound for the absolute value of the continued fraction.*

Proof. By (1.3) we have $B_2=1+a_2\neq 0$, $B_3=1+a_2+a_3\neq 0$; and if we

* Bulletin of the American Mathematical Society, abstract 45-7-277.

† Worpitzky, Jahresbericht, Friedrichs-Gymnasium und Realschule, Berlin, 1865, pp. 3–39. Independent proofs of this theorem were given later by Pringsheim and by Van Vleck.

put $c_n = a_{n+1}B_{n-1}/B_{n+1}$, $n=1, 2, 3, \cdots$, we have $|c_1| = |a_2/(1+a_2)| \leqq r_1$, $|c_2| = |a_3/(1+a_2+a_3)| \leqq r_2$. Let us assume that $B_{n+1} \neq 0$, $|c_n| \leqq r_n$, for $n=1, 2, 3, \cdots, k$, $k \geqq 2$, and prove that the same holds for $n=k+1$, and hence for all n. Inasmuch as

$$\left| \frac{A_{n+1}}{B_{n+1}} - \frac{A_n}{B_n} \right| = |\, c_1 c_2 \cdots c_n \,|,$$

it will follow that $1+\sum r_1 r_2 \cdots r_n$ is a majorant for the series (1.2), so that the continued fraction converges if $1+\sum r_1 r_2 \cdots r_n$ converges. Since $B_n \neq 0$, for all n, the continued fraction converges if some a_n is 0.

From the recursion relation $B_n = B_{n-1} + a_n B_{n-2}$, $n \geqq 2$, we easily obtain the formula

$$B_{k+2} = (1 + a_{k+1} + a_{k+2})B_k - a_k a_{k+1} B_{k-2}, \tag{2.1}$$

$k=2, 3, 4, \cdots$. If $a_{k+2} \neq 0$, so that $r_{k+1} > 0$, then we have

$$\frac{B_{k+2}}{a_{k+2}B_k} = \frac{1 + a_{k+1} + a_{k+2}}{a_{k+2}} - \frac{a_{k+1}}{a_{k+2}} \frac{a_k B_{k-2}}{B_k},$$

or, by (1.3) and our assumption,

$$\left| \frac{B_{k+2}}{a_{k+2}B_k} \right| \geqq \left| \frac{1 + a_{k+1} + a_{k+2}}{a_{k+2}} \right| - \left| \frac{a_{k+1}}{a_{k+2}} \right| r_{k-1} \geqq \frac{1}{r_{k+1}} > 0.$$

Hence $B_{k+2} \neq 0$, and $|c_{k+1}| \leqq r_{k+1}$. On the other hand, if $a_{k+2} = 0$, it follows from the fundamental recursion formula that $B_{k+2} = B_{k+1} \neq 0$, and $|c_{k+1}| = 0 \leqq r_{k+1}$. This completes the induction and the proof of Theorem A.

Since Theorem A is simply a comparison theorem for the series (1.2), we have at once by the *Weierstrass "M-test,"* the following theorem on *uniform convergence* of continued fractions.

THEOREM 2.1. *If in the continued fraction* (1.1) *of Theorem* A, *the elements* a_n *are functions of any variables over a domain D in which the inequalities* (1.3) *hold, with the* r_n *independent of the variables, and if the series* $1+\sum r_1 r_2 \cdots r_n$ *converges, then the continued fraction converges uniformly over D.*

If in Theorem A we use the series

$$\frac{A_{k+1}}{B_{k+1}} + \left(\frac{A_{k+2}}{B_{k+2}} - \frac{A_{k+1}}{B_{k+1}} \right) + \left(\frac{A_{k+3}}{B_{k+3}} - \frac{A_{k+2}}{B_{k+2}} \right) + \cdots \tag{2.2}$$

instead of the series (1.2) we get by the same methods the more general result:

THEOREM A′. *If there exist an index* $k \geqq 0$, *and numbers* $r_n \geqq 0$, $n = k+1, k+2, k+3, \cdots$, *such that* $A_{k+1}, B_{k+2}, B_{k+3} \neq 0$ *and*

$$r_{k+1} \left| B_{k+2} A_{k+1} \right| \geqq \left| a_2 a_3 \cdots a_{k+2} \right|, \qquad r_{k+2} \left| B_{k+3} \right| \geqq \left| a_{k+3} B_{k+1} \right|,$$
$$r_n \left| 1 + a_n + a_{n+1} \right| \geqq r_n r_{n-2} \left| a_n \right| + \left| a_{n+2} \right|,$$

$n = k+3, k+4, k+5, \cdots$, *then the series*

$$\left| \frac{A_{k+1}}{B_{k+1}} \right| \left(1 + \sum_{n=k+1}^{\infty} r_{k+1} r_{k+2} \cdots r_n \right) \tag{2.3}$$

is a majorant for the series (2.2). *The continued fraction converges if some* a_n *is* 0 *for* $n \geqq k+2$, *or if the series* (2.3) *converges.*

3. **Relation of Theorem A to other convergence criteria.** We now prove a second theorem.

THEOREM B. *Let* $p_1, p_2, p_3, \cdots$ *be real and positive,* $p_1 > 1$. *Then the continued fraction* (1.1) *in which the elements* a_n *are functions of any variables converges uniformly over the domain characterized by the inequalities*

$$\left| a_n \right| \leqq (p_n - 1)/p_n p_{n-1}, \qquad n = 2, 3, 4, \cdots. \tag{3.1}$$

The quantity

$$1 + \frac{p_1}{p_1 - 1} \left\{ \frac{1}{p_1} - \frac{1}{1 + \sum_1^{\infty} (p_1 - 1)(p_2 - 1) \cdots (p_n - 1)} \right\} \tag{3.2}$$

is an upper bound for the absolute value of the continued fraction over the domain, and is actually attained if $a_n = (1 - p_n)/p_n p_{n-1}$, $n = 2, 3, 4, \cdots$.

This theorem includes the *general Pringsheim criterion.*† However, our theorem goes farther and gives an upper bound for the absolute value of the continued fraction. The greatest improvement is in respect to the question of uniform convergence. *Pringsheim required that the series* $\sum (p_1 - 1)(p_2 - 1) \cdots (p_n - 1)$ *be divergent in order to obtain uniform convergence of the continued fraction.*‡

Proof of Theorem B. Let $t_n = (p_n - 1) p_n p_{n-1}$, and form the continued fraction

$$1/1 - t_2/1 - t_3/1 - \cdots. \tag{3.3}$$

Let G_n/H_n be the nth convergent of (3.3) and put

$$H_n^* = p_1 p_2 \cdots p_n H_n, \quad n = 1, 2, 3, \cdots, \qquad H_0^* = 1,$$

† Perron, p. 258, Theorem 26. It is to be noted that there is no loss in generality in assuming $p_1 > 1$ in that theorem.

‡ Perron, p. 262, Theorem 30.

so that $H_n^* = p_n H_{n-1}^* - p_n p_{n-1} t_n H_{n-2}^*$, $n = 2, 3, 4, \cdots$, and

$$H_n^* - H_{n-1}^* = (p_n - 1)(H_{n-1}^* - H_{n-2}^*) = (p_1 - 1)(p_2 - 1) \cdots (p_n - 1),$$

$n = 1, 2, 3, \cdots$. Thus $H_n^* \geqq 1$, and consequently $H_n > 0$ for all n.

We take for r_n in Theorem A the quantity

$$r_n = t_{n+1} H_{n-1}/H_{n+1}, \qquad n = 1, 2, 3, \cdots.$$

Inasmuch as $H_{n+1} = (1 - t_n - t_{n+1}) H_{n-1} - t_n t_{n-1} H_{n-3}$, we readily verify that $r_n(1 - t_n - t_{n+1}) = r_n r_{n-2} t_n + t_{n+1}$, $n = 3, 4, 5, \cdots$, $(r_0 = r_{-1} = t_1 = 0)$. Hence it follows from (3.1) and the definition of t_n that the elements of the continued fraction satisfy the inequalities of Theorem A with the values of the r_n which we have determined.

Since

$$r_1 r_2 \cdots r_n = \frac{p_1}{p_1 - 1}\left[\frac{1}{H_n^*} - \frac{1}{H_{n+1}^*}\right],$$

it follows that the series $1 + \sum r_1 r_2 \cdots r_n$ converges; and since the r_n are independent of the variables, it follows by Theorem 2.1 that the continued fraction converges uniformly over the domain of the variables.

The quantity (3.2) is simply the sum of the majorant series, and is therefore an upper bound for the absolute value of the continued fraction. It is attained by the continued fraction when $a_n = -t_n$.

This completes the proof of Theorem B.

By specializing the p_n in Theorem B the familiar *special Pringsheim criteria*† may be obtained, with the attendant upper bounds. For example, setting $p_n = 2$ we obtain this theorem:

THEOREM 3.1. (Worpitzky.) *The continued fraction* (1.1) *converges if* $|a_n| \leqq 1/4$, $n = 2, 3, 4, \cdots$.

In this case the upper bound for the absolute value of the continued fraction is 2, and if $a_n = -1/4$, this value is attained.

The importance of Theorem B is illustrated by the fact that the proof of the following theorem of Van Vleck as given by Perron‡ can now be considerably shortened.

THEOREM 3.2. (Van Vleck.) *Let* $s_1, s_2, s_3, \cdots$ *be real numbers, and* $t_1, t_2, t_3, \cdots$ *complex numbers, such that*

$$0 < s_1 < 1, \quad 0 \leqq s_n < 1, \qquad n = 2, 3, 4, \cdots,$$
$$|t_n| \leqq 1, \qquad n = 1, 2, 3, \cdots;$$

† Perron, pp. 259–260.

‡ Perron, pp. 262–264.

and let x be a complex variable. Then the continued fraction

$$\frac{s_1 t_1 x^{n_1}}{1} + \frac{s_2(1-s_1)t_2 x^{n_2}}{1} + \frac{s_3(1-s_2)t_3 x^{n_3}}{1} + \cdots$$

(n_i positive integers) converges uniformly for $|x| \leqq 1$. The modulus of the analytic function which the continued fraction represents cannot exceed

$$s_1 |t_1 x^{n_2}| \left(1 + \frac{1-s_1}{s_1}\left[1 - \frac{1}{1 + (1-s_1)\sum_1^\infty \frac{s_1 s_2 \cdots s_n}{(1-s_1)(1-s_2)\cdots(1-s_n)}}\right]\right).$$

Proof. Take $p_n = 1/(1-s_n)$, $n = 1, 2, 3, \cdots$, in Theorem B.

As an illustration of the last theorem, the function $x/(\arctan x)$ has the continued fraction expansion*

$$1 + \frac{x^2}{1} + \frac{\frac{1\cdot 3}{4x^2}}{1} + \frac{\frac{3\cdot 5}{9x^2}}{1} + \frac{\frac{5\cdot 7}{}}{1} + \cdots,$$

to which Theorem 3.2 is applicable with $s_n = n/(2n+1)$. We find that the modulus of the function does not exceed $1+|x^2|$ if $|x| \leqq 1$.

In the proof of the next theorem we use for the first time the *second* interpretation of the fundamental inequalities, which was mentioned in the introduction.

THEOREM C. *The continued fraction of Theorem* A *converges at least in the wider sense if*

$$\liminf_{n=\infty} (r_1 r_2 r_3 \cdots r_n) = 0, \tag{3.4}$$

and converges to a finite value if (3.4) *holds and actual inequality holds at least once in* (1.3).

Proof. We need only consider the case where $a_n \neq 0$, $n = 2, 3, 4, \cdots$, since Theorem A disposes of the case where some a_n is 0. If $a_n \neq 0$ for all n, it is clear that $r_n > 0$ for all n.

Consider the continued fractions

$$\frac{1}{1+a_2} - \frac{a_2 a_3}{1+a_3+a_4} - \frac{a_4 a_5}{1+a_5+a_6} - \cdots, \tag{3.5}$$

$$1 - \frac{a_2}{1+a_2+a_3} - \frac{a_3 a_4}{1+a_4+a_5} - \frac{a_5 a_6}{1+a_6+a_7} - \cdots. \tag{3.6}$$

* Perron, p. 351, formula (17) (in slightly different form).

The approximants of (3.5) are the *even* approximants* of (1.1), while those of (3.6) are the *odd* approximants of (1.1). We now write the inequalities (1.3) in the form

$$\frac{r_1}{|a_2|}|1+a_2| \geqq 1,$$

$$\frac{r_{2n+1}}{|a_{2n+2}|}|1+a_{2n+1}+a_{2n+2}| \geqq \frac{r_{2n+1}}{|a_{2n+2}|}\frac{r_{2n-1}}{|a_{2n}|}|a_{2n}a_{2n+1}|+1; \tag{3.7}$$

$$\frac{r_2}{|a_3|}|1+a_2+a_3| \geqq 1,$$

$$\frac{r_{2n+2}}{|a_{2n+3}|}|1+a_{2n+2}+a_{2n+3}| \geqq \frac{r_{2n+2}}{|a_{2n+3}|}\frac{r_{2n}}{|a_{2n+1}|}|a_{2n+1}a_{2n+2}|+1, \tag{3.8}$$

$n=1, 2, 3, \cdots$. *But the conditions* (3.7) *constitute a Pringsheim test for* (3.5) *provided inequality holds at least once*;† and the same is true for conditions (3.8) and the continued fraction (3.6). It should be noted that the continued fractions (3.5), (3.6) may diverge to ∞ if at every stage in the conditions (3.7), (3.8) equality holds. In any case the limits $\lim_{n=\infty}(A_{2n}/B_{2n})=L_0$, $\lim_{n=\infty}(A_{2n-1}/B_{2n-1})=L_1$ exist, finite or infinite.

But by Theorem A,

$$\left|\frac{A_{n+1}}{B_{n+1}}-\frac{A_n}{B_n}\right| \leqq r_1r_2\cdots r_n$$

so that by (3.4), $L_0=L_1$. If inequality holds at least once in (1.3), then this common limit of the two sequences is finite, and the continued fraction converges.

In a recent paper, Leighton‡ obtained a convergence theorem for (1.1) by applying the Pringsheim inequalities to (3.7) and (3.8) separately. Since the element a_1 appearing in his work can have no bearing upon the convergence of the continued fraction, it will be seen that his result may be stated as follows:

THEOREM 3.3. (Leighton.) *Let*§

$$|1+a_2| \geqq 1, \qquad |1+a_2+a_3| \geqq 1,$$
$$|1+a_n+a_{n+1}| \geqq |a_{n-1}a_n|+1, \tag{3.9}$$

* Perron, p. 201, formulas (7), (8).

† See the second footnote.

‡ Walter Leighton, Duke Mathematical Journal, vol. 4 (1938), pp. 775–778.

§ The proof as given by Leighton would not permit equality in the first two relations, which, however, we allow only when some a_n vanishes.

$n=3,4,5,\cdots$. Then the continued fraction (1.1) *converges if some a_n vanishes, or if actual inequality holds in the first two relations* (3.9) *and* $\liminf |a_n| < \infty$.

Proof. Multiply the inequalities (3.9) by $|a_2|, |a_3|, |a_4|, \cdots$, respectively, and they take the form (1.3) where $r_n = |a_{n+1}|$. Hence if some a_n vanishes, the continued fraction converges by Theorem A.

If $a_n \neq 0$, $n \geqq 2$, then by hypothesis actual inequality holds in the first two relations (1.3), so that (3.5), (3.6) converge. Moreover, the quantities

$$g_1 = |B_2| - \frac{|a_2|}{r_1}|B_0|, \qquad g_2 = |B_3| - \frac{|a_3|}{r_2}|B_1|$$

are *positive*. We then have by (2.1), (1.3), the inequalities

$$\left\{|B_{k+2}| - \frac{|a_{k+2}|}{r_{k+1}}|B_k|\right\} \geqq r_{k-1}|a_{k+1}|\left\{|B_k| - \frac{|a_k|}{r_{k-1}}|B_{k-2}|\right\}, \qquad k \geqq 2,$$

and therefore

$$\begin{aligned} |B_{2n+2}| - \frac{|a_{2n+2}|}{r_{2n+1}}|B_{2n}| &\geqq r_1 r_3 \cdots r_{2n-1}|a_3 a_5 \cdots a_{2n+1}|\, g_1 > 0, \\ |B_{2n+3}| - \frac{|a_{2n+3}|}{r_{2n+2}}|B_{2n+1}| &\geqq r_2 r_4 \cdots r_{2n}|a_4 a_6 \cdots a_{2n+2}|\, g_2 > 0. \end{aligned} \tag{3.10}$$

When $r_n = |a_{n+1}|$, these inequalities become

$$\begin{aligned} |B_{2n+2}| - |B_{2n}| &\geqq |a_2 a_3 \cdots a_{2n+1}|\, g_1, \\ |B_{2n+3}| - |B_{2n+1}| &\geqq |a_3 a_4 \cdots a_{2n+2}|\, g_2. \end{aligned} \tag{3.11}$$

Consequently,

$$\begin{aligned} \left|\frac{A_{2n+2}}{B_{2n+2}} - \frac{A_{2n+1}}{B_{2n+1}}\right| &\leqq \frac{|a_2|}{g_2}\,\frac{1}{|B_{2n+2}|}\left(\left|\frac{B_{2n+3}}{B_{2n+1}}\right| - 1\right), \\ \left|\frac{A_{2n+1}}{B_{2n+1}} - \frac{A_{2n}}{B_{2n}}\right| &\leqq \frac{1}{g_1}\,\frac{1}{|B_{2n+1}|}\left(\left|\frac{B_{2n+2}}{B_{2n}}\right| - 1\right). \end{aligned} \tag{3.12}$$

Since by (3.11) the sequences $|B_{2n}|$ and $|B_{2n-1}|$ are strictly monotone increasing, they must have limits, finite or infinite. If either has a finite limit, we see from (3.12), remembering that (3.5), (3.6) both converge, that the continued fraction (1.1) converges. If $\lim |B_{2n}| = \lim |B_{2n-1}| = \infty$, we have the inequalities

$$\left|\frac{B_{2n+2}}{B_{2n}B_{2n+1}}\right| \leqq \frac{1}{|B_{2n}|} + \frac{|a_{2n+2}|}{|B_{2n+1}|}, \qquad \left|\frac{B_{2n+3}}{B_{2n+1}B_{2n+2}}\right| \leqq \frac{1}{|B_{2n+1}|} + \frac{|a_{2n+3}|}{|B_{2n+2}|},$$

so that under the hypothesis lim inf $|a_n| < \infty$, the right-hand member of at least one of (3.12) must have the inferior limit 0, and, therefore, the continued fraction converges in this case also.

Of course, the inequalities (3.10) hold whenever the $r_n > 0$ exist satisfying (1.3) with inequality for $n = 1, 2$. We therefore have

$$\left|\frac{A_{2n+3}}{B_{2n+3}} - \frac{A_{2n+2}}{B_{2n+2}}\right| = \frac{|a_2 a_3 \cdots a_{2n+3}|}{|B_{2n+2}B_{2n+3}|}$$

$$\leqq \frac{|a_{2n+3}|}{g_1 g_2 r_1 r_2 \cdots r_{2n}} \left\{1 - \frac{|a_{2n+2}|}{r_{2n+1}}\left|\frac{B_{2n}}{B_{2n+2}}\right|\right\}\left\{1 - \frac{|a_{2n+3}|}{r_{2n+2}}\left|\frac{B_{2n+1}}{B_{2n+3}}\right|\right\}$$

$$< \frac{|a_{2n+3}|}{g_1 g_2 r_1 r_2 \cdots r_{2n}},$$

and, similarly,

$$\left|\frac{A_{2n+2}}{B_{2n+2}} - \frac{A_{2n+1}}{B_{2n+1}}\right| < \frac{|a_{2n+2}|}{g_1 g_2 r_1 r_2 \cdots r_{2n-1}}.$$

Now if lim inf $(r_1 r_2 \cdots r_n) = 0$, (1.1) converges by Theorem C. If $r_1 r_2 \cdots r_n$ is bounded away from 0, these inequalities show that if lim inf $|a_n| = 0$ the continued fraction converges. Hence we have proved

THEOREM 3.4. *If in the continued fraction* (1.1) *the inequalities* (1.3) *hold with inequality for* $n = 1, 2$, *then the continued fraction converges if* lim inf $|a_n| = 0$.

4. **Convergence theorems involving** $\sum |b_n|$. If the elements a_n are different from 0, the continued fraction (1.1) can be thrown into the form $1/b_1 + 1/b_2 + 1/b_3 + \cdots$, where $b_1 = 1$, $a_n = 1/b_n b_{n-1}$, $n = 2, 3, 4, \cdots$. Therefore, by a theorem of von Koch,* *the even and odd approximants of the continued fraction will have distinct limits, finite or infinite, if the series* $\sum |b_n|$ *converges. Then a necessary condition for the convergence of the continued fraction is the divergence of the series* $\sum |b_n|$.

In this section we shall give some conditions under which the divergence of the series $\sum |b_n|$ is *sufficient* to insure the convergence of the continued fraction.

THEOREM D. *If there exist positive numbers* r_n *satisfying the inequalities* (1.3), *with actual inequality holding for at least one even and one odd index, and if*

$$r_1 r_3 r_5 \cdots r_{2n-1} < M, \qquad r_2 r_4 r_6 \cdots r_{2n} < M, \tag{4.1}$$

* Perron, p. 235, Theorem 6.

$n=1, 2, 3, \cdots$, where M is a finite constant, then a necessary and sufficient condition for the convergence of the continued fraction

$$1/1 + a_2/1 + a_3/1 + a_4/1 + \cdots, \qquad a_n \neq 0, \tag{4.2}$$

is that the series $\sum |b_n|$ diverge, where $b_1=1$, $a_n=1/b_{n-1}b_n$, $n=2, 3, 4, \cdots$.

Proof. We may assume that actual inequality holds in (1.3) for $n=1, 2$. For if inequality holds for a certain index n, we may slightly increase $r_{n-2}, r_{n-4}, \cdots$ in such a way as to introduce inequality for $n=1, 2$. This cannot affect condition (4.1) since the values of only a finite number of the r_n have been changed.

In terms of the b_n the continued fraction (4.2) takes the form

$$1/b_1 + 1/b_2 + 1/b_3 + \cdots. \tag{4.3}$$

If P_n/Q_n is the nth approximant of (4.3) one may easily verify the relation

$$b_{n+1}Q_{n+3} = (b_{n+1}b_{n+2}b_{n+3} + b_{n+1} + b_{n+3})Q_{n+1} - b_{n+3}Q_{n-1}, \tag{4.4}$$

$n=1, 2, 3, \cdots$. From this we obtain by means of (1.3) the inequality

$$|Q_{n+3}| - \frac{1}{r_{n+2}}|Q_{n+1}| \geqq r_n\left|\frac{b_{n+3}}{b_{n+1}}\right|\left(|Q_{n+1}| - \frac{1}{r_n}|Q_{n-1}|\right), \tag{4.5}$$

$n=1, 2, 3, \cdots$. Since (1.3) holds with actual inequality for $n=1, 2$, the quantities

$$e_1 = \frac{|Q_2|r_1 - |Q_0|}{|b_2|r_1}, \qquad e_2 = \frac{|Q_3|r_2 - |Q_1|}{|b_3|r_2}$$

are positive. On introducing the proper factors in (4.5) and summing we get

$$r_3r_5\cdots r_{2n+1}|Q_{2n+2}| \geqq e_1\left(1 + \sum_{k=1}^{n} r_1r_3^2r_5^2\cdots r_{2k-1}^2r_{2k+1}|b_{2k+2}|\right) \tag{4.6}$$

and

$$r_4r_6\cdots r_{2n+2}|Q_{2n+3}| \geqq e_2\left(1 + \sum_{k=1}^{n+1} r_2r_4^2r_6^2\cdots r_{2k-2}^2r_{2k}|b_{2k+1}|\right). \tag{4.7}$$

Now

$$\left|\frac{P_n}{Q_n} - \frac{P_{n+1}}{Q_{n+1}}\right| = \frac{1}{|Q_nQ_{n+1}|}; \tag{4.8}$$

and the limits $\lim_{n=\infty}(P_{2n}/Q_{2n})$, $\lim_{n=\infty}(P_{2n-1}/Q_{2n-1})$ both exist and are finite. *Hence the continued fraction will converge if $1/|Q_nQ_{n+1}|$ has the lower limit* 0. By Theorem C, if $r_1r_2\cdots r_n$ has the lower limit 0, our theorem is

granted. Hence there remains only the case where, for a constant $c>0$,

$$c \leqq r_3 r_5 \cdots r_{2n-1} < M, \qquad c \leqq r_4 r_6 \cdots r_{2n} < M.$$

The theorem now follows from a consideration of (4.6), (4.7), and the hypothesis that the series $\sum |b_n|$ is divergent.

THEOREM E. *If the numbers a_n in* (1.1) *are all less in absolute value than a constant M, and if there exist positive numbers r_n satisfying the inequalities* (1.3) *with actual inequality for an even and an odd index, then the continued fraction converges.*

Proof. We may suppose the $a_n \neq 0$, for this case is treated in Theorem A. Furthermore we may suppose that $\liminf (r_1 r_2 \cdots r_n) \geqq c > 0$, for otherwise Theorem C applies. From a composition of corresponding terms in the inequalities (4.6), (4.7) we get

$$r_3 r_4 \cdots r_{n+1} \left| Q_{n+1} Q_{n+2} \right| \geqq e_1 e_2 \left(1 + \sum r_1 r_2 r_3^2 r_4^2 \cdots r_{n-1}^2 r_n r_{n+1} \left| b_{n+1} b_{n+2} \right| \right).$$

If $\limsup (r_1 r_2 \cdots r_n) \leqq N$, where N is finite, then

$$\limsup \left| Q_{n+1} Q_{n+2} \right| \geqq \frac{e_1 e_2}{N} \left(r_1 r_2 + \sum_{k=1}^{\infty} r_1^2 r_2^2 \cdots r_{k-1}^2 r_k r_{k+1} M \right),$$

and it is evident that the series on the right diverges. Then we have $\limsup |Q_{n+1} Q_{n+2}| = \infty$. If N is infinite we have $|Q_{n+1} Q_{n+2}| > e_1 e_2 (r_1 r_2 \cdots r_{n+1}) M$, and $\limsup |Q_{n+1} Q_{n+2}| = \infty$. The theorem now follows from (4.8).

5. **Convergence regions.** By a *convergence region* for the continued fraction (1.1) we shall understand a set W of points in the complex plane such that if $a_2, a_3, a_4, \cdots$ are arbitrary points in W the continued fraction converges. A convergence region is necessarily *bounded*. For if W were unbounded, we could choose the $a_n \neq 0$ in W in such a way that the series $\sum |b_n|$ would be convergent, where $b_1 = 1$, $a_n = 1/b_{n-1} b_n$, and for this choice of the a_n the continued fraction would diverge by oscillation. On the other hand, if the $a_n \neq 0$ lie in any bounded region, it is clear that the series $\sum |b_n|$ must *diverge*.

Worpitzky* showed that the interior and boundary of the circle $|z| = 1/4$ is a convergence region for (1.1); and since (1.1) diverges if $a_n = -1/4 - c$ where c is real and positive, it follows that the interior and boundary of a circle with center at the origin and radius greater than 1/4 cannot be a convergence region. Szász† obtained convergence regions containing the origin, which neither contain nor are contained in the circle $|z| = 1/4$.

* See the second footnote on page 156.

† Szász, Journal für die reine und angewandte Mathematik, vol. 147 (1916), pp. 132–160.

We now propose to characterize completely all convergence regions which are symmetric with respect to the real axis.

In order to obtain a *necessary restriction* upon such a region, consider the continued fraction $1/1+\bar{z}/1+z/1+\bar{z}/1+z/1+\cdots$, in which $z=x+iy$, $\bar{z}=x-iy$, x, y real. If this continued fraction converges, then the continued fraction

$$-z+\frac{z}{1+}\,\frac{\bar{z}}{1+}\,\frac{z}{1+}\,\frac{\bar{z}}{1+}\,\frac{z}{1+}\cdots \tag{5.1}$$

must converge, as well as the continued fraction

$$\frac{|z^2|}{1+2x}\,-\,\frac{|z^2|}{1+2x}\,-\,\frac{|z^2|}{1+2x}\,-\cdots,$$

having as approximants the odd approximants of (5.1). Now if the latter converges, it must have a real value, which is a root of the quadratic equation in u:

$$u^2-(1+2x)u+(x^2+y^2)=0.$$

Consequently, it is *necessary* for the convergence of (5.1) that $y^2\leqq x+1/4$, that is, z shall lie within or upon the parabola $|z|-\Re(z)=1/2$, where $\Re(z)$ denotes the real part of z. *Therefore a convergence region for* (1.1) *which is symmetric with respect to the real axis is necessarily a finite region bounded by this parabola.* We shall prove that this necessary condition is also a sufficient condition.

THEOREM F. (The parabola theorem.) *If the elements a_n of the continued fraction* (1.1) *lie within or upon the parabola* (*see figure*)

$$|z|-\Re(z)=1/2, \tag{5.2}$$

then:

(a) *The denominator of the nth approximant A_n/B_n is different from* 0 *for all n.*

(b) *The sequences of even and odd approximants have finite limits L_0 and L_1, and $|L_0-L_1|\leqq 1$.*

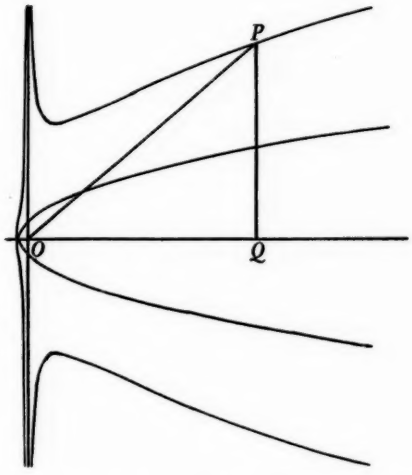
(c) *If $a_n\neq 0$, $n=2, 3, 4, \cdots$, the continued fraction converges if and only if the series $\sum|b_n|$ diverges, where $b_1=1$, $a_n=1/b_nb_{n-1}$, $n=2, 3, 4, \cdots$.*

(d) *If some a_n vanishes, the continued fraction converges and equals one of its approximants.*

(e) *The parabola* (5.2) *is the best possible curve symmetric with respect to the real axis having these properties.*

Proof. Part (e) follows from the fact shown above that when $a_{2n}=\bar{z}$, $a_{2n-1}=z$, the continued fraction diverges if z lies outside the parabola.

The other parts of the theorem will follow from Theorems A and D, if we show that the inequalities (1.3) hold for $r_n=1$, with actual inequality for $n=1, 2$.



Put $a_n=u_n+iv_n$ where u_n, v_n are real. Then if a_n lies in or upon the parabola (5.2), we must have

$$|a_n| = u_n + h_n/2, \qquad n = 2, 3, 4, \cdots,$$

where $0 \leqq h_n \leqq 1$. We then have

$$\begin{aligned}
|1 + a_2| &\geqq 1 + u_2 > u_2 + h_2/2 = |a_2|,\\
|1 + a_2 + a_3| &\geqq 1 + u_2 + u_3 = (1 - h_2/2 - h_3/2) + |a_2| + |a_3| > |a_3|,\\
|1 + a_n + a_{n+1}| &\geqq 1 + u_n + u_{n+1} \geqq u_n + h_n/2 + u_{n+1} + h_{n+1}/2\\
&\geqq |a_n| + |a_{n+1}|,
\end{aligned}$$

$n=3, 4, 5, \cdots$, as was to be proved.

It is of interest to note the analogy between the above result and the following important and general convergence criterion of Van Vleck:*

THEOREM 5.1. (Van Vleck.) *Let* $b_n=|b_n|e^{i\theta_n}$ *where*

$$-\pi/2 + \epsilon \leqq \theta_n \leqq \pi/2 - \epsilon, \qquad \epsilon > 0,$$

* Perron, pp. 264–271.

$n=1, 2, 3, \cdots$, and suppose not all the b_n with odd index are 0. The then sequences of even and odd approximants of the continued fraction

$$1/b_1 + 1/b_2 + 1/b_3 + \cdots$$

have finite limits; and the continued fraction converges if and only if the series $\sum|b_n|$ diverges.

It is interesting to note that the expression $|a_n| - \Re(a_n)$ appearing in the parabola theorem occurs in the following theorem of Szász:*

THEOREM 5.2. (Szász.) *The continued fraction* (1.1) *converges if the series $\sum|a_n|$ converges, and*

$$\sum_2^\infty \{|a_n| - \Re(a_n)\} < 2.$$

It is easy to construct examples where convergence can be established by Theorem 5.1 but not by Theorem A or our later theorems, and vice versa. The same statement applies to Theorem 5.2, where, however, at most *three* of the a_n's can lie outside the parabola. But if, for example, a_2 is taken outside the parabola equal to $-1+\epsilon$, and $a_3=a_4=1$, then if the other a_n's are chosen so that

$$\sum_5^\infty |a_n| - \sum_5^\infty \Re(a_n) < 2\epsilon, \qquad \epsilon > 0,$$

the conditions of Szász are satisfied, but the inequalities (1.3) cannot hold for nonnegative r_n. In fact, we must have $r_1 \geqq |a_2|/|1+a_2| = 1/\epsilon - 1$. But $|1+a_3+a_4| < r_1|a_3|$ if $\epsilon < 1/4$, and hence it is not possible to determine $r_3 \geqq 0$.

We have seen that a convergence region for (1.1) which is symmetric with respect to the real axis must lie in the parabola (5.2). We now turn our attention to the problem of finding convergence regions which are not symmetric with respect to the real axis.

THEOREM G. (The triangle theorem.) *Let $P(x, y)$ be any point upon the curve (see figure after Theorem* F)

$$y = (2x + 1)(4x + 1)^{1/2}/2x, \tag{5.3}$$

in the plane of $z=x+iy$. Let O denote the origin and Q the point $z=x$. Then the closed region made up of the interior and boundary of the triangle OPQ is a convergence region for the continued fraction (1.1).

Proof. As in the proof of Theorem F, we show that Theorem D applies

* Perron, p. 259.

with $r_n=1$. Let m be the slope of the line OP. Then if a_n is in the triangle OPQ, we may write $a_n=u_n(1+imt_n)$, where u_n is the real part of a_n, and $0\leqq t_n\leqq 1$. Since $u_2>-1/2$ it follows that $|1+a_2|>|a_2|$, so that the first inequality (1.3) holds with $r_1=1$. Since $u_n\geqq -1/4$, $u_2u_3\geqq 0$, $v_2v_3\geqq 0$, where v_ni is the imaginary part of a_n, it is easily seen that $|1+a_2+a_3|>|a_3|$, so that the second inequality (1.3) holds. Thus actual inequality holds in the first two relations (1.3). The remaining inequalities will hold provided

$$(5.4)\qquad \begin{aligned}&(1+2u_n+2u_{n+1}+2u_nu_{n+1})\\ &\qquad\geqq 2u_nu_{n+1}\left[(1+m^2s^2)^{1/2}(1+m^2t^2)^{1/2}-2m^2s^2t^2\right],\end{aligned}$$

$n=3, 4, 5, \cdots$, where $s=t_n$, $t=t_{n+1}$. We find that the maximum of the bracketed part in the right-hand member of (5.4) for s, t between 0 and 1 is $(1+m^2)^{1/2}$. Hence (5.4) will hold if

$$(5.5)\qquad \left[(1+m^2)^{1/2}-1\right]u_nu_{n+1}-u_n-u_{n+1}\leqq 1/2.$$

We next maximize the left-hand member of (5.5) in the two cases $u_n>0$, $u_n<0$, and find that in the first case (5.5) will hold if $w=u_n$ satisfies the inequality

$$(5.6)\qquad w\leqq \left[2\left\{(1+\left[1+m^2\right]^{1/2})\right\}^{1/2}-2\right]^{-1},$$

while in the second case (5.5) will hold provided $w=-u_n$ satisfies (5.6). But this is precisely the condition that the point P shall lie upon the curve (5.3). The proof of Theorem G is now complete.

It is easy to show that *any bounded set of points lying upon a ray from the origin, in the right half-plane, is a convergence region for* (1.1). In fact, upon such a ray the inequalities (1.3) hold for $r_n=1$ with actual inequality throughout. To see this, put $a_n=\alpha_ne^{2i\theta_n}$ in (1.3), and the latter take the form

$$(5.7)\qquad \begin{aligned}&(1+\alpha_n+\alpha_{n+1})^2-\left(r_{n-2}\alpha_n+\frac{\alpha_{n+1}}{r_n}\right)^2\\ &\qquad\geqq 4\left\{\alpha_n\sin^2\theta_n+\alpha_n\alpha_{n+1}\sin^2(\theta_{n+1}-\theta_n)+\alpha_{n+1}\sin^2\theta_{n+1}\right\},\end{aligned}$$

$n=1, 2, 3, \cdots$, $(r_0=r_{-1}=\alpha_1=0)$. When $r_n=1$, $\theta_n=\theta=$const., these reduce to

$$1+2\alpha_2\geqq 0,\qquad (\alpha_2+\alpha_3)\cos 2\theta\geqq -(1+\alpha_2^2)/(\alpha_2+\alpha_3),$$
$$(\alpha_n+\alpha_{n+1})\cos 2\theta\geqq -1/2,$$

$n=3, 4, 5, \cdots$. These are obviously satisfied with actual inequality throughout, if $-\pi/2\leqq 2\theta\leqq\pi/2$.

As an application, consider an arbitrary sequence $x_1, x_2, x_3, \cdots$ with nonnegative real parts. Let v_n be the ray from the origin upon which x_n lies.

Choose y_n upon v_n but in the parabola (5.2). Then if $a_2=x_1$, $a_3=y_1$, $a_4=y_2$, $a_5=x_2$, $a_6=y_2$, $a_7=y_3$, $a_8=x_3$, $\cdots$, it is clear that the inequalities (1.3) are satisfied with $r_n=1$. If, in particular, the y_n have a limit point in the finite plane (and they can evidently be so chosen), then the continued fraction $1/1+a_2/1+a_3/1+\cdots$ converges by Theorem D. *This shows that it is possible for the a_n to be everywhere dense in the right half-plane for a convergent continued fraction* (1.1).*

6. **The problem of Szász.**† Szász proposed the following general problem: *To assign neighborhoods to the elements of a given convergent continued fraction in which these elements may vary without destroying the convergence property of the continued fraction.* We shall indicate in this section how Theorem A may be applied to this problem.

We consider a convergent continued fraction

$$\frac{1}{1}+\frac{c_2}{1}+\frac{c_3}{1}+\cdots \tag{6.1}$$

and seek to determine positive numbers $t_2, t_3, t_4, \cdots$ such that if

$$|a_n - c_n| \leqq t_n, \qquad n = 2, 3, 4, \cdots, \tag{6.2}$$

then the continued fraction

$$\frac{1}{1}+\frac{a_2}{1}+\frac{a_3}{1}+\cdots$$

will converge. The theorems of the preceding sections may be applied to this problem in certain cases. For that purpose put

$$a_n = c_n + \delta_n t_n, \qquad n = 2, 3, 4, \cdots, 0 \leqq |\delta_n| \leqq 1,$$

in the inequalities (1.3). Then these inequalities will hold provided

$$\begin{aligned} r_1|1+c_2| &\geqq |c_2| + (1+r_1)t_2, \\ r_2|1+c_2+c_3| &\geqq |c_3| + (1+r_2)t_3 + r_2 t_2, \\ r_n|1+c_n+c_{n+1}| &\geqq r_n r_{n-2}|c_n| + |c_{n+1}| + (1+r_n)t_{n+1} + r_n(1+r_{n-2})t_n, \end{aligned} \tag{6.3}$$

$n=3, 4, 5, \cdots$. It follows that if there exist positive numbers r_n, t_n satisfying (6.3), then

$$r_n|1+c_n+c_{n+1}| \geqq r_n r_{n-2}|c_n| + |c_{n+1}|, \tag{6.4}$$

$n=1, 2, 3, \cdots$, $(r_{-1}=r_0=a_1=0)$, and hence under suitable restrictions, (1.1)

* Leighton and Wall showed that this can be done for the entire plane. See American Journal of Mathematics, vol. 58 (1936), pp. 267–281; p. 269.

† See the second footnote on page 165.

will converge if (6.2) holds. If the a_n, c_n, t_n are functions of variables, and the r_n are independent of the variables, then under suitable conditions the convergence will be *uniform*.

It is convenient to specialize the inequalities (6.3) in the following way. Put

$$s_1 = \frac{r_1|1+c_2|-|c_2|}{1+r_1}, \qquad s_2 = \frac{r_2|1+c_2+c_3|-|c_3|}{1+2r_2}$$

$$s_n = \frac{r_n|1+c_n+c_{n+1}|-r_nr_{n-2}|c_n|-|c_{n+1}|}{1+2r_n+r_nr_{n-2}}, \tag{6.5}$$

$n=3, 4, 5, \cdots$. If $s_1, s_2, s_3, \cdots \geqq 0$, and t_n' is the larger of s_{n-1}, s_n, then $t_n=t_n'$, $(n\geqq 2)$, satisfy the inequalities (6.3). We therefore have this theorem:

THEOREM 6.1. *Let $a_2, a_3, a_4, \cdots, c_2, c_3, c_4, \cdots$ be functions of any variables over a domain D, and suppose that there exist numbers $r_1, r_2, r_3, \cdots$, nonnegative and independent of the variables such that the quantities s_n, defined by* (6.5), *are nonnegative over D. Then if*

$$|a_n - c_n| \leqq \begin{cases} s_{n-1} \\ s_n \end{cases}, \qquad n = 2, 3, 4, \cdots, \tag{6.6}$$

we shall have $r_n|1+a_n+a_{n+1}| \geqq r_nr_{n-2}|a_n|+|a_{n+1}|$ for $n=1, 2, 3, \cdots$, $(r_{-1}=r_0=a_1=0)$. Hence, in particular, if $1+\sum r_1r_2\cdots r_n$ converges, the continued fraction (1.1) *will converge uniformly over D.*

As an illustration, let $c_n=(2n-3)/4$, $n\geqq 2$; $r_n=n/(n+2)$, $n\geqq 1$. Then $s_1=1/8$, $s_n=(n-1)/(4n)$, $n\geqq 2$, and hence the continued fraction (1.1) in which $a_2=(1+4x_2)/4$, $a_3=(3+4x_3)/4, \cdots, a_n=(2n-3+4x_n)/4, \cdots$ converges uniformly for

$$|x_2| \leqq 1/8, \qquad |x_n| \leqq 1/4 - 1/[4(n-1)],$$

$n=3, 4, 5, \cdots$.

If $c_n=c$ and $\Re(c)>-1/4$, we can determine by means of Theorem 6.1 a neighborhood of c which is a convergence region for (1.1).

THEOREM 6.2. *Let $c, a_2, a_3, a_4, \cdots$ be functions of any variables over a certain domain D in which*

$$|a_n - c| \leqq \frac{|1+2c|-2|c|}{4}, \qquad n = 2, 3, 4, \cdots.$$

Then the continued fraction (1.1) *converges uniformly over D.*

Proof. Take $r_n=1$, $c_n=c$ in Theorem 6.1. Then the theorem follows at once by application of Theorem D.

It is not difficult to show by means of Theorem A, with $r_n=n/(n+2)$, that 2 is the least upper bound of the absolute value of the continued fraction in Theorem 6.2.

Theorem 6.2 gives a generalization of the Worpitzky theorem, which reduces to the latter when $c=0$. It is applicable whenever the real part of c exceeds $-1/4$, and in this case furnishes a solution of the problem of Szász for the periodic continued fraction $1/1+c/1+c/1+\cdots$. Szász obtained a solution except when c is real and not greater than $-1/4$, in which case there is no solution. We note that when $c=1$, the theorem of Szász gives as the radius of the convergence circle approximately 5/22, Theorem 6.1 gives the value 1/4, and the *parabola theorem* gives the value 1. Moreover, from the latter theorem it follows that the radius of a convergence circle with center $c=1$ cannot exceed 1.

ARMOUR INSTITUTE OF TECHNOLOGY,
CHICAGO, ILL.,
NORTHWESTERN UNIVERSITY,
EVANSTON, ILL.

HARMONIC MINIMAL SURFACES*

BY

W. C. GRAUSTEIN

1. **Introduction.** A minimal surface in a Euclidean space of three dimensions is harmonic if it is representable in terms of Cartesian coordinates (x_1, x_2, x_3) by an equation of the form $U(x_1, x_2, x_3) = \text{const.}$, where U is a harmonic function. The problem of the determination of all harmonic minimal surfaces is equivalent to a problem in hydrodynamics which will be described in §2. Hamel† has recently solved this problem in two ways. His first method demands in itself that the functions under consideration be real. On the other hand, the second method places no restriction on the functions involved and, since it appears to lead to the same results as the first, one is tempted to infer that all solutions $U(x_1, x_2, x_3)$ of the problem are real. Actually, there exist imaginary solutions and they are geometrically far more intriguing than the real solutions.

It is shown in this paper that all solutions, real and imaginary, except those with isotropic gradients, are reducible, by means of a change of coordinate axes and an integral linear transformation on the function itself, to one of the following forms:

(I) $$U = \tan^{-1}(x_2/x_3) + ax_1,$$

(IIa) $$(2zU + ix_1)^2 = z \tan (4zU^2 + 4ix_1U + \bar{z}),$$

(IIb) $$z^4U^6 + 3(x_1^2 + x_2^2 + x_3^2) = 0,$$

(IIIa) $$U = -\int \frac{dy}{(1 - y^3)^{1/2}} = \frac{2}{3}\int \frac{du}{(1 + u^2)^{5/6}},$$

$$(z^2 - 2ix_1)^3u^2 + (z^3 - 3ix_1z + \tfrac{3}{2}\bar{z})^2 = 0,$$

(IV) $$U = f(z)x_1 + \phi(z),$$

where $z = x_2 + ix_3$ and $\bar{z} = x_2 - ix_3$ throughout. It is to be noted that only in the last case does the solution involve arbitrary functions.

The families of minimal surfaces $U = \text{const.}$ in the five cases are: (I) a family of helicoids or a pencil of planes with Euclidean axis; (IIa) a family of imaginary transcendental surfaces; (IIb) a family of imaginary quartic surfaces; (IIIa) a family of imaginary sextic surfaces; and (IV) families of

* Presented to the Society, December 29, 1939; received by the editors December 2, 1939.

† *Potentialströmungen mit konstanter Geschwindigkeit*, Sitzungsberichte der Preussischen Akademie der Wissenschaften, 1937, pp. 5–20.

BOSTON UNIVERSITY
COLLEGE OF LIBERAL ARTS
LIBRARY

imaginary cylinders with isotropic rulings, or a pencil of planes with isotropic axis, or a pencil of parallel planes.

The helicoids of a family (I) are all congruent and each of them admits a one-parameter group of screw motions about a Euclidean axis into itself. The transcendental surfaces (IIa) are all congruent under a one-parameter group of "rotations" about an isotropic line. Each of the sextic surfaces (IIIa) admits a "screw motion" about a line at infinity into itself. Finally, the quartic surfaces of the family (IIb) are all congruent and each admits a one-parameter group of "rotations" about an isotropic line into itself. Furthermore, this family of surfaces belongs to a triply orthogonal system of surfaces which admits a two-parameter group of rigid motions into itself.

It is of interest to note that in every case, not merely those just cited, a one-parameter group of rigid motions plays an important role, and it is perhaps still more striking that these groups exhaust all the one-parameter groups of complex rigid motions.

Analytically, the problem calls for the simultaneous solutions of two partial differential equations of the second order in three independent variables. A frontal attack on it from this point of view seems hopeless. In fact, no matter how it is approached, the analytic complications are severe. The method here adopted turned out to be the same as the second method employed by Hamel. It makes use of the intrinsic geometry of surfaces and congruences of curves. In particular, it introduces three mutually orthogonal congruences of curves, with unit tangent vector fields α, β, γ, which are closely associated with the required family of surfaces $U = \text{const.}$, and expresses the desired properties of these surfaces by a suitable choice of the coefficients in the equations of variations of α, β, γ with respect to the arcs of the curves of the three congruences. These equations of variation constitute the differential system finally to be integrated. Their conditions of integrability yield a second differential system of ten partial differential equations of the first order in five dependent and three independent variables. The analytic difficulties lie primarily in the solution of this second system and they are not rendered any easier by the fact that the independent variables are the nonholonomic arcs of the curves of the three congruences.

The paper falls into five parts. In Part A, the general case of the problem is formulated after the manner just outlined, and the solutions of the secondary or scalar differential system are listed. In Part B, the primary or vector differential system is integrated in the various cases and the functions U found, and in Part C, the properties of the corresponding minimal surfaces $U = \text{const.}$ are discussed. The deductions of the solutions of the scalar differential system and the proof that there are no other solutions is given in Part

D, and Part E is devoted to a special case, previously excluded, which gives rise to the solution (IV).

It is assumed that all functions are analytic in the complex variables x_1, x_2, x_3.

A. Formulation of problem and method of solution

2. **The physical problem.** Let there be given a stationary irrotational flow of a frictionless incompressible liquid with the special property that the velocity along an arbitrary line of flow is constant along this line.

A flow of the general type described is characterized by the existence of a harmonic function $U(x_1, x_2, x_3)$ whose gradient is the flow-vector. It will have the desired special property provided the gradient of the velocity of flow, or of any variable function of the velocity, is orthogonal always to the flow-vector. Thus, the problem of determining all flows of the kind required is identical with the problem of finding the simultaneous solutions of the two partial differential equations

$$\Delta_2 U = 0, \qquad \Delta_1(U, V) = 0, \tag{1}$$

where

$$V = \log (\Delta_1 U)^{1/2}. \tag{2}$$

Equivalent to these equations are the relations $\Delta_2 U = 0, \Delta_2 U - \Delta_1(U, V) = 0$ which characterize the function U as harmonic and the surfaces $U = \text{const.}$ as minimal.

In the general case in which $\Delta_1 V \neq 0$, we prefer to employ the following equations:

$$\Delta_2 U - \Delta_1(U, V) = 0, \qquad \Delta_1(U, V) = 0. \tag{3}$$

Thus, our problem becomes that of determining the families of minimal surfaces $U = \text{const.}$ which are cut orthogonally by the corresponding families of surfaces $V = \text{const.}$, where V and U are related by conditions (2). It is in this form that we shall solve the general problem in the complex domain.

The special case in which $\Delta_1 V = 0$ will be treated in §17.

3. **Geometric formulation of the analytic problem.** If U is a solution of (3) for which $\Delta_1 V \neq 0$, the families of surfaces $U = \text{const.}$ and $V = \text{const.}$ are mutually orthogonal and determine three mutually orthogonal congruences of curves: the orthogonal trajectories C_1 of the surfaces $U = \text{const.}$, the orthogonal trajectories C_2 of the surfaces $V = \text{const.}$, and the curves of intersection C_3 of the two families of surfaces. The curves of these congruences, properly directed, have respectively the unit tangent vectors

$$(4)\qquad \alpha = \frac{\nabla U}{(\Delta_1 U)^{1/2}}, \qquad \beta = \frac{\nabla V}{(\Delta_1 V)^{1/2}}, \qquad \gamma = \widetilde{\alpha\beta},$$

where ∇U, for example, is the gradient of U and γ is the vector product of α and β.

The curves C_2 and C_3 lying on a generic surface S of the family $U=\text{const.}$ form an orthogonal system. If $1/r_2$, $1/\tau_2$, $1/\rho_2$ are the normal curvature, geodesic torsion, and geodesic curvature of the directed curves C_2, with respect to $+\alpha$ as the unit vector normal to S, and $1/r_3$, $1/\tau_3$, $1/\rho_3$ have the same meanings for the directed curves C_3, then $1/\tau_2+1/\tau_3=0$ and, since S is minimal, $1/r_2+1/r_3=0$.

From (4) and (2) it follows that the differential of arc ds_1 of the curves C_1 has the value $e^{-V}dU$. Hence, the curves C_1 are geodesics on the surfaces S' of the family $V=\text{const.}$, and β and γ play for them the roles of principal normal vector and binomial vector, respectively. Furthermore, the torsion $1/T_1$ of the curves C_1 is equal to the geodesic torsion of these curves, as curves on the surfaces S'. This geodesic torsion is the negative of the geodesic torsion of the curves C_3, as curves on the surfaces S' or, since the surfaces S and S' intersect under a constant angle, as curves on the surfaces S, and hence it is equal to $1/\tau_2$. Thus, $1/T_1=1/\tau_2$.

By means of these results we obtain from the Frenet-Serret formulas for the curves C_1, and from the formulas for the variation of the surface trihedrals* of the curves C_2 and C_3, as curves on the surfaces S, the following equations:

$$(5)\qquad \begin{aligned}
&\frac{\partial\alpha}{\partial s_1} = A\beta, && \frac{\partial\alpha}{\partial s_2} = B\beta + C\gamma, && \frac{\partial\alpha}{\partial s_3} = C\beta - B\gamma,\\
&\frac{\partial\beta}{\partial s_1} = -A\alpha - C\gamma, && \frac{\partial\beta}{\partial s_2} = -B\alpha + E\gamma, && \frac{\partial\beta}{\partial s_3} = -C\alpha + F\gamma,\\
&\frac{\partial\gamma}{\partial s_1} = C\beta, && \frac{\partial\gamma}{\partial s_2} = -C\alpha - E\beta, && \frac{\partial\gamma}{\partial s_3} = B\alpha - F\beta,
\end{aligned}$$

where $\partial/\partial s_1$, $\partial/\partial s_2$, $\partial/\partial s_3$ represent directional differentiation in the positive directions of the curves C_1, C_2, C_3 respectively, and

$$(6)\qquad A = \frac{1}{R_1}, \quad B = \frac{1}{r_3} = -\frac{1}{r_2}, \quad C = \frac{1}{T_1} = \frac{1}{\tau_2} = -\frac{1}{\tau_3}, \quad E = \frac{1}{\rho_2}, \quad F = \frac{1}{\rho_3},$$

$1/R_1$ being, of course, the curvature of the curves C_1.

Inasmuch as, for an arbitrary function $f(x_1, x_2, x_3)$,

* See Graustein, *Differential Geometry*, p. 165.

$$\frac{\partial f}{\partial s_1} = \sum \alpha_i \frac{\partial f}{\partial x_i}, \qquad \frac{\partial f}{\partial s_2} = \sum \beta_i \frac{\partial f}{\partial x_i}, \qquad \frac{\partial f}{\partial s_3} = \sum \gamma_i \frac{\partial}{\partial x_i},$$

it follows, by (5), that

$$\begin{aligned} \frac{\partial}{\partial s_2}\frac{\partial f}{\partial s_1} - \frac{\partial}{\partial s_1}\frac{\partial f}{\partial s_2} &= A\frac{\partial f}{\partial s_1} + B\frac{\partial f}{\partial s_2} + 2C\frac{\partial f}{\partial s_3}, \\ \frac{\partial}{\partial s_3}\frac{\partial f}{\partial s_2} - \frac{\partial}{\partial s_2}\frac{\partial f}{\partial s_3} &= E\frac{\partial f}{\partial s_2} + F\frac{\partial f}{\partial s_3}, \\ \frac{\partial}{\partial s_1}\frac{\partial f}{\partial s_3} - \frac{\partial}{\partial s_3}\frac{\partial f}{\partial s_1} &= B\frac{\partial f}{\partial s_3}. \end{aligned} \tag{7}$$

These relations we shall refer to as the conditions of integrability $(f;\ s_1,\ s_2)$, $(f;\ s_2,\ s_3)$, $(f;\ s_3,\ s_1)$, respectively.

From (4) we have, in view of (2),

$$\frac{\partial U}{\partial s_1} = e^V, \qquad \frac{\partial U}{\partial s_2} = 0, \qquad \frac{\partial U}{\partial s_3} = 0, \tag{8}$$

$$\frac{\partial V}{\partial s_1} = 0, \qquad \frac{\partial V}{\partial s_2} = A, \qquad \frac{\partial V}{\partial s_3} = 0, \tag{9}$$

where

$$A = (\Delta_1 V)^{1/2} \neq 0. \tag{10}$$

The conditions of integrability of equations (8) simply require that the quantity A in (9) and (10) be identical with the quantity A in equations (5) to (7).

The conditions of integrability of equations (9) are

$$\frac{\partial A}{\partial s_1} = -AB, \qquad \frac{\partial A}{\partial s_3} = AE. \tag{11}$$

Equations (5) and (11), together with the inequality (10), constitute necessary conditions. Suppose, conversely, that the scalar functions A $(\neq 0)$, B, C, E, F and the vector functions α, β, γ, representing three mutually orthogonal unit vector fields with the same disposition as the coordinate axes, are known solutions of equations (5) and (11). Equations (9), for the given α, β, γ, and A, are then integrable and the function V is determined to within an additive constant; and equations (8), for the given α, β, γ and the V just found, are integrable and U is determined to within a multiplicative and an additive constant.

Since (8) and (9) are equivalent to the relations $\nabla U = e^V \alpha$, $\nabla V = A\beta$, it follows that $V = \log (\Delta_1 U)^{1/2}$, $\Delta_1 V = A^2$, and $\Delta_1(U, V) = 0$. Hence, the function V is related to U as prescribed by (2), $\Delta_1 V \neq 0$, and the families of surfaces $U = \text{const.}$ and $V = \text{const.}$ cut orthogonally. Finally, since α is a unit vector normal to the surfaces $U = \text{const.}$ and the curves C_2 and the curves C_3 determined respectively by the unit vector fields β and γ lie on these surfaces, it follows from (5) that the surfaces are minimal.

Thus, we have established the following existence theorem.

THEOREM 1. *A necessary and sufficient condition that there exist a harmonic function $U(x_1, x_2, x_3)$ such that the surfaces $U = \text{const.}$ are minimal is that there exist three mutually orthogonal unit vector fields α, β, γ, with the same disposition as the axes, and five scalar functions A $(\neq 0)$, B, C, E, F which satisfy equations* (5) *and* (11). *The function U is then determined to within a multiplicative and an additive constant and can be found by quadratures.*

COROLLARY. *If two solutions α, β, γ, A $(\neq 0)$, B, C, E, F of equations* (5) *and* (11) *are related to one another by a rigid motion, the two resulting families of surfaces $U = \text{const.}$ are congruent.*

The corollary is an obvious consequence of the fact that the equations with which we are dealing are invariant with respect to the group of rigid motions.

4. **Outline of the solution.** The nine conditions of integrability of equations (5), combined with the two equations in (11), yield the following ten independent equations in A, B, C, E, F and their directional derivatives:

$$\text{(12a)} \qquad \frac{\partial A}{\partial s_1} = -AB, \qquad \frac{\partial A}{\partial s_2} = A^2 + 2B^2 + 2C^2 - AF, \qquad \frac{\partial A}{\partial s_3} = AE,$$

$$\text{(12b)} \qquad \begin{aligned} \frac{\partial B}{\partial s_1} &= B^2 - C^2 - AF, & \frac{\partial B}{\partial s_3} - \frac{\partial C}{\partial s_2} &= 2CF + 2BE, \\ \frac{\partial C}{\partial s_1} &= AE + 2BC, & \frac{\partial B}{\partial s_2} + \frac{\partial C}{\partial s_3} &= 2CE - 2BF, \end{aligned}$$

$$\text{(12c)} \qquad \begin{aligned} \frac{\partial E}{\partial s_1} + \frac{\partial C}{\partial s_2} &= -BE - 2CF, \\ \frac{\partial F}{\partial s_1} + \frac{\partial C}{\partial s_3} &= AB + BF, \\ \frac{\partial E}{\partial s_3} - \frac{\partial F}{\partial s_2} &= -B^2 - C^2 + E^2 + F^2. \end{aligned}$$

The process of solving our problem now becomes clearer. First, the system of equations (12) is to be solved for the unknown functions A, B, C, E, F. For the values found for these functions, equations (5) will be completely integrable and will determine, to within a rigid motion, three mutually perpendicular unit vector fields α, β, γ, and equations (9) and (8) will, then, yield the desired functions V and U. This, at least, would be the general procedure if, instead of the directional derivatives, we had ordinary partial derivatives with which to deal. Actually, equations (12) involve the unknown vector functions α, β, γ (through the directional derivatives), as well as the unknown scalars A, B, C, E, F, and equations (5) involve α, β, γ both in the derivatives and as the unknown functions. Nevertheless, the general procedure described remains valid, as we shall proceed to show.

The essential requirement for the employment of this procedure is that the three mutually orthogonal unit vector fields α, β, γ and the directional derivatives in the directions of α, β, γ which are employed when equations (12) are solved for A, B, C, E, F should later be found to satisfy equations (5) for the values of A, B, C, E, F obtained. This requirement is actually fulfilled by the inherent demand that the directional derivatives in question enjoy the conditions of integrability (7). For, it is readily proved that, if (7) are satisfied, the vector fields α, β, γ and the derivatives in their directions satisfy (5).

Equations (12), subject to the integrability conditions (7), have the following solutions:

$$
\text{(I)}\qquad
\begin{array}{lll}
B = 0, & E = 0, & AF + C^2 = 0, \\
\dfrac{\partial A}{\partial s_1} = 0, & \dfrac{\partial A}{\partial s_2} = A^2 + 3C^2, & \dfrac{\partial A}{\partial s_3} = 0, \\
\dfrac{\partial C}{\partial s_1} = 0, & \dfrac{\partial C}{\partial s_2} = -2CF, & \dfrac{\partial C}{\partial s_3} = 0, \\
\dfrac{\partial F}{\partial s_1} = 0, & \dfrac{\partial F}{\partial s_2} = C^2 - F^2, & \dfrac{\partial F}{\partial s_3} = 0,
\end{array}
$$

$$
\text{(II)}\qquad
\begin{array}{lll}
AE = -2BC, & AF = 2B^2, & A^2 + 4(B^2 + C^2) = 0, \\
\dfrac{\partial A}{\partial s_1} = -AB, & \dfrac{\partial A}{\partial s_2} = \tfrac{1}{2}A^2 - 2B^2, & \dfrac{\partial A}{\partial s_3} = -2BC, \\
\dfrac{\partial B}{\partial s_1} = \tfrac{1}{4}A^2, & \dfrac{\partial B}{\partial s_2} = AB, & \dfrac{\partial B}{\partial s_3} = \tfrac{1}{2}AC, \\
\dfrac{\partial C}{\partial s_1} = 0, & \dfrac{\partial C}{\partial s_2} = \tfrac{1}{2}AC, & \dfrac{\partial C}{\partial s_3} = 0,
\end{array}
$$

$$E = 0, \qquad F = -\tfrac{1}{2}A, \qquad A^2 + 4(B^2 + C)^2 = 0,$$

$$\frac{\partial A}{\partial s_1} = -AB, \qquad \frac{\partial A}{\partial s_2} = A^2, \qquad \frac{\partial A}{\partial s_3} = 0,$$

(III)
$$\frac{\partial B}{\partial s_1} = \tfrac{1}{2}A^2 + B^2 - C^2, \qquad \frac{\partial B}{\partial s_2} = AB, \qquad \frac{\partial B}{\partial s_3} = 0,$$

$$\frac{\partial C}{\partial s_1} = 2BC, \qquad \frac{\partial C}{\partial s_2} = AC, \qquad \frac{\partial C}{\partial s_3} = 0.$$

In all three cases, $A \neq 0$. Henceforth, this condition will always be tactily understood.

Solutions (II) and (III) are imaginary, whereas (I) exists in the real domain. It may be readily verified that all three satisfy equations (12) and the integrability conditions (7).

The derivation of the three solutions from (12) and the proof that (12) has no other solutions present analytic problems of unusual complexity. In order not to interrupt the present development, we shall postpone the consideration of them to Part D.

Since, for the values of A, B, C, E, F furnished by a solution of (12), equations (5), (9), and (8) are completely integrable, it is theoretically possible to solve equations (5) for the three mutually orthogonal unit vector functions α, β, γ and hence (9) and (8) for the scalar functions V and U. Practically, however, this procedure is complicated by the presence of directional, rather than partial, derivatives, and we find ourselves forced to adopt a different method.

We remark, first, that equations (8) and (9) are of the same type as the differential equations in one of the solutions of (12), and hence that U and V are just as much known as the quantities involved in these differential equations. Consequently, we have at our disposal the seven functions U, V, A, B, C, E, F. It is evident from (I), (II), (III) that at most two of the last five are functionally independent.

By integration of equations (5) it is possible to find α, β, γ in terms of certain of the seven functions. For these values of α, β, γ, the equations

$$\frac{\partial x}{\partial s_1} = \alpha, \qquad \frac{\partial x}{\partial s_2} = \beta, \qquad \frac{\partial x}{\partial s_3} = \gamma \tag{13}$$

are completely integrable, by virtue of (5), and x_1, x_2, x_3 may be found in terms of three independent functions, or parameters, one of which is U. Elimination of the other two parameters from the three equations results in the desired value of U as a function of x_1, x_2, x_3.

Geometrically, each of the three solutions (I), (II), (III) has two cases according as $C \neq 0$ or $C = 0$, that is, by (6), according as the curves C are twisted or plane. The two cases we shall distinguish by attaching the letters a and b to the Roman numerical. It is readily verified that the solutions (IIb) and (IIIb) are identical.

B. The harmonic functions

5. **The real solutions.** *Case* Ia. Inasmuch as $B = E = 0$, we conclude from (6) that the curves C_2 are straight lines. Since $C \neq 0$, the Gaussian curvature* of the surfaces S is not zero. Hence, the surfaces S are right helicoids and a normal form for the function U is

$$U = \tan^{-1}(x_2/x_3) + ax_1 = 0, \qquad a \neq 0. \tag{14}$$

The lines of flow in the physical problem are the circular helices cutting the helicoids orthogonally.

Case Ib. It may be shown geometrically that the surfaces S form a pencil of intersecting planes and that a normal form for U is (14), where $a = 0$. The lines of flow are circles.

It will be advantageous to illustrate the analytic method described at the end of the preceding section in this simple case. Since $B = C = E = F = 0$, equations (I) reduce to

$$\frac{\partial A}{\partial s_1} = 0, \qquad \frac{\partial A}{\partial s_2} = A^2, \qquad \frac{\partial A}{\partial s_3} = 0. \tag{15}$$

Comparison of these equations with (9) shows that we may take $V = \log A$. Then, (8) becomes

$$\frac{\partial U}{\partial s_1} = A, \qquad \frac{\partial U}{\partial s_2} = 0, \qquad \frac{\partial U}{\partial s_3} = 0. \tag{16}$$

Equations (5) reduce to $\partial\alpha/\partial s_1 = A\beta$, $\partial\beta/\partial s_1 = -A\alpha$, with the remaining derivatives of α and β, and all of those of γ, zero. Consequently, in view of (16), if a, b, c are three fixed mutually perpendicular unit vectors with the disposition of the axes, we have

$$\alpha = b\cos U - c\sin U, \qquad \beta = -b\sin U - c\cos U, \qquad \gamma = -a.$$

But then equations (13), since a function W exists satisfying the equations

$$\frac{\partial W}{\partial s_1} = 0, \qquad \frac{\partial W}{\partial s_2} = 0, \qquad \frac{\partial W}{\partial s_3} = -1, \tag{17}$$

* For the formula employed, see Graustein, *Invariant methods in classical differential geometry*, Bulletin of American Mathematical Society, vol. 36 (1930), p. 508.

have the integral

$$x = Wa + e^{-V}(b \sin U + c \cos U) + k,$$

where k represents a triple of constants.

By means of a rigid motion this representation is reducible to the normal form

$$(18) \qquad x_1 = W, \qquad x_2 = e^{-V} \sin U, \qquad x_3 = e^{-V} \cos U.$$

Eliminating V, we obtain for U the normal form $\tan^{-1}(x_2/x_3)$.

Equations (18) represent a change from Cartesian coordinates to curvilinear coordinates (U, V, W). In view of equations (15), (16), (17) and the fact that $A = e^V$, the congruences of parametric curves consist of the curves C_1, the curves C_2, and the curves C_3, respectively, and the parametric surfaces form a triply orthogonal system, that of cylindrical coordinates.

6. **The imaginary solution IIa.** The point of departure here is solution (II) of equations (12) for $C \neq 0$, namely,

$$AE = -2BC, \quad AF = 2B^2, \quad A^2 + 4(B^2 + C^2) = 0, \quad C \neq 0,$$

$$(19) \qquad \begin{aligned} &\frac{\partial A}{\partial s_1} = -AB, && \frac{\partial A}{\partial s_2} = \tfrac{1}{2}A^2 - 2B^2, && \frac{\partial A}{\partial s_3} = -2BC, \\ &\frac{\partial B}{\partial s_1} = \tfrac{1}{4}A^2, && \frac{\partial B}{\partial s_2} = AB, && \frac{\partial B}{\partial s_3} = \tfrac{1}{2}AC, \\ &\frac{\partial C}{\partial s_1} = 0, && \frac{\partial C}{\partial s_2} = \tfrac{1}{2}AC, && \frac{\partial C}{\partial s_3} = 0. \end{aligned}$$

Equations (5) become

$$(20) \qquad \begin{aligned} &\frac{\partial \alpha}{\partial s_1} = A\beta, && \frac{\partial \alpha}{\partial s_2} = B\beta + C\gamma, && \frac{\partial \alpha}{\partial s_3} = C\beta - B\gamma, \\ &\frac{\partial \beta}{\partial s_1} = -A\alpha - C\gamma, && \frac{\partial \beta}{\partial s_2} = -B\alpha - \frac{2BC}{A}\gamma, && \frac{\partial \beta}{\partial s_3} = -C\alpha + \frac{2B^2}{A}\gamma, \\ &\frac{\partial \gamma}{\partial s_1} = C\beta, && \frac{\partial \gamma}{\partial s_2} = -C\alpha + \frac{2BC}{A}\beta, && \frac{\partial \gamma}{\partial s_3} = B\alpha - \frac{2B^2}{A}\beta. \end{aligned}$$

It follows from (19) and (20) that the determinant of β and its first two derivatives with respect to s_2 vanishes. Hence, the curves C_2 are plane curves. The planes in which they lie are determined by the vector fields β and $B\alpha - E\gamma$.

Similarly, it can be shown that the curves C_3 are plane curves, lying in the planes determined by the vector fields γ and $B\alpha - F\beta$.

A vector field common to the planes of C_2 and C_3 is $B\alpha - F\beta - E\gamma$ or, since $B \neq 0$ by (19), $A\alpha - 2B\beta + 2C\gamma$. The vectors of this field are isotropic. Moreover, they are fixed in direction. In fact, it is readily shown, by (19) and (20), that the derivatives of the vector field

$$a = \frac{A^2}{C^3}\alpha - \frac{2AB}{C^3}\beta + \frac{2AC}{C^3}\gamma \tag{21}$$

all vanish.

A vector field normal to the planes of the curves C_3 is $2B\alpha + A\beta$. Setting

$$\eta = \frac{2B}{C}\alpha + \frac{A}{C}\beta, \tag{22}$$

we find that

$$\frac{\partial \eta}{\partial s_1} = -\tfrac{1}{2}C^2 a, \qquad \frac{\partial \eta}{\partial s_2} = 0, \qquad \frac{\partial \eta}{\partial s_3} = 0.$$

Comparison of (9) with the derivatives of C in (19) shows that $e^V = kC^2$. Since k becomes the multiplicative constant in the value of U, we may without loss of generality specialize it. Taking $k = -1/2$, and noting that (8) then becomes

$$\frac{\partial U}{\partial s_1} = -\tfrac{1}{2}C^2, \qquad \frac{\partial U}{\partial s_2} = 0, \qquad \frac{\partial U}{\partial s_3} = 0, \tag{23}$$

we conclude that

$$\eta = Ua + b, \tag{24}$$

where b is a triple of constants.

Equation (21) and the equation obtained by equating the values of η in (22) and (24) can be solved for β and γ as linear combinations of α, a, b. Substituting these values of β and γ in the equations for the derivatives of α in (20), we find that the resulting equations can be written in the forms:

$$\frac{\partial}{\partial s_1}\left(\frac{C^3}{A^2}\alpha\right) = \frac{C^4}{A^2}Ua + \frac{C^4}{A^2}b,$$

$$\frac{\partial}{\partial s_2}\left(\frac{C^3}{A^2}\alpha\right) = \left(\frac{C^6}{2A^3} + \frac{2BC^4}{A^3}U\right)a + \frac{2BC^4}{A^3}b,$$

$$\frac{\partial}{\partial s_3}\left(\frac{C^3}{A^2}\alpha\right) = \left(-\frac{BC^5}{2A^3} + \frac{C^5 - B^2C^3}{A^3}U\right)a + \frac{C^5 - B^2C^3}{A^3}b.$$

It may be shown, by means of (19) and (23), that the coefficients of b in

these three equations are respectively the directional derivatives of the function $U/4-BC^2/2A^2$ and that $-BC^4/2A^2$, $C^6/2A^3$, $-BC^5/2A^3$ are the derivatives of the function $-C^4/8A^2$. Hence, the equations are integrable and yield the relation

$$\frac{C^3}{A^2}\alpha = \left(\frac{1}{8}U^2 - \frac{BC^2}{2A^2}U - \frac{C^4}{8A^2}\right)a + \left(\frac{1}{4}U - \frac{BC^2}{2A^2}\right)b + c, \tag{25}$$

where c is a triple of constants.

Having solved equations (20) for α, β, γ, we could proceed by the method outlined at the end of §4 to find x_1, x_2, x_3. We adopt a different, but equivalent, method. Instead of solving equations (21), (22), (24), (25) for α, β, γ in terms of a, b, c, we solve for a, b, c in terms of α, β, γ. Making use of the function $H=4B/C^3$, whose directional derivatives are the coefficients of α, β, γ in (21), we find the expressions

$$\begin{aligned}
a &= \frac{\partial H}{\partial s_1}\alpha + \frac{\partial H}{\partial s_2}\beta + \frac{\partial H}{\partial s_3}\gamma, \\
b &= \left(-\frac{\partial H}{\partial s_1}U + \frac{2B}{C}\right)\alpha + \left(-\frac{\partial H}{\partial s_2}U + \frac{A}{C}\right)\beta - \frac{\partial H}{\partial s_3}U\gamma, \\
c &= \left(\frac{1}{8}\frac{\partial H}{\partial s_1}U^2 - \frac{B}{2C}U - \frac{C}{8}\right)\alpha + \left(\frac{1}{8}\frac{\partial H}{\partial s_2}U^2 - \frac{A}{4C}U + \frac{BC}{4A}\right)\beta \\
&\quad + \left(\frac{1}{8}\frac{\partial H}{\partial s_3}U^2 + \frac{C^2}{4A}\right)\gamma.
\end{aligned} \tag{26}$$

The conditions on a, b, c guaranteeing that α, β, γ are mutually perpendicular unit vector fields with the same disposition as the axes are now readily found to be*

$$\begin{aligned}
&(a\,|\,a) = 0, \quad (b\,|\,b) = -4, \quad (c\,|\,c) = 0, \\
&(b\,|\,c) = 0, \quad (c\,|\,a) = 1, \quad (a\,|\,b) = 0,
\end{aligned} \qquad (a\ b\ c) = -2. \tag{27}$$

It is to be noted that the vector c, as well as the vector a, is isotropic.

The integrals of equations (26) are

$$\begin{aligned}
&(a\,|\,x) + a_0 = H, \qquad H = 4B/C^3, \\
&(b\,|\,x) + b_0 = -HU - 2/C, \\
&(c\,|\,x) + c_0 = \tfrac{1}{8}HU^2 + \tfrac{1}{2}(1/C)\,U + \tfrac{1}{8}\tan^{-1}(C/B),
\end{aligned} \tag{28}$$

* By $(a\,|\,b)$ is meant the scalar product of the vectors a and b, and by $(a\ b\ c)$, the determinant of the components of three vectors a, b, c.

for equations (26) simply state that a, b, c are respectively the gradients of the functions which appear on the right-hand sides of (28).

Equations (28) define x_1, x_2, x_3 as functions of U, B, C—three parameters which by (19) and (23) are independent. Denoting the linear functions of x_1, x_2, x_3 on the left-hand sides by $a(x), b(x), c(x)$ respectively, and eliminating B and C, we obtain the equation

$$[a(x)U + b(x)]^2 = a(x) \tan [a(x)U^2 + 2b(x)U + 8c(x)], \tag{29}$$

which with the attendant conditions (27) serves to define, within a multiplicative and an additive constant, the general solution U in this case.

It follows from (27) that the equations

$$a(x) = 4(x_2' + ix_3'), \qquad b(x) = 2ix_1', \qquad 8c(x) = x_2' - ix_3'$$

define a rigid motion. By means of this rigid motion or, what is the same thing, by setting a_0, b_0, c_0 equal to zero and taking as a, b, c respectively the triples

$$0, 4, 4i, \qquad 2i, 0, 0, \qquad 0, \tfrac{1}{8}, -\tfrac{1}{8}i, \tag{30}$$

(29) reduces to the normal form

$$(2zU + ix_1)^2 = z \tan (4zU^2 + 4ix_1U + \bar{z}), \tag{29a}$$

where $z = x_2 + ix_3$ and $\bar{z} = x_2 - ix_3$. Hence, there is essentially only one solution U of our problem in this case.

7. **The imaginary solution** IIb. When $C = 0$, equations (II) become

$$E = 0, \quad F = -\tfrac{1}{2}A, \quad A^2 + 4B^2 = 0, \quad C = 0, \quad B \neq 0,$$

$$\begin{aligned} &\frac{\partial A}{\partial s_1} = -AB, && \frac{\partial A}{\partial s_2} = A^2, && \frac{\partial A}{\partial s_3} = 0, \\ &\frac{\partial B}{\partial s_1} = -B^2, && \frac{\partial B}{\partial s_2} = AB, && \frac{\partial B}{\partial s_3} = 0. \end{aligned} \tag{31}$$

Hence, equations (5) reduce to

$$\begin{aligned} &\frac{\partial \alpha}{\partial s_1} = A\beta, && \frac{\partial \alpha}{\partial s_2} = B\beta, && \frac{\partial \alpha}{\partial s_3} = -B\gamma. \\ &\frac{\partial \beta}{\partial s_1} = -A\alpha, && \frac{\partial \beta}{\partial s_2} = -B\alpha, && \frac{\partial \beta}{\partial s_3} = -\tfrac{1}{2}A\gamma. \\ &\frac{\partial \gamma}{\partial s_1} = 0, && \frac{\partial \gamma}{\partial s_2} = 0, && \frac{\partial \gamma}{\partial s_3} = B\alpha + \tfrac{1}{2}A\beta. \end{aligned} \tag{32}$$

The curves C_3 turn out to be plane curves, lying in the isotropic planes determined by the vector field γ and the isotropic vector field $B\alpha+\frac{1}{2}A\beta$. The vectors of the latter field are fixed in direction. In fact, we find that the vector

$$a = BK\alpha + \tfrac{1}{2}AK\beta, \tag{33}$$

where K ($\neq 0$) is a function defined by the compatible equations

$$\frac{\partial K}{\partial s_1} = -BK, \qquad \frac{\partial K}{\partial s_2} = -\tfrac{1}{2}AK, \qquad \frac{\partial K}{\partial s_3} = 0, \tag{34}$$

is a fixed vector.

It follows from (32) that

$$\gamma = -Wa + b, \tag{35}$$

where W is defined by the integrable equations

$$\frac{\partial W}{\partial s_1} = 0, \qquad \frac{\partial W}{\partial s_2} = 0, \qquad \frac{\partial W}{\partial s_3} = -\frac{1}{K}, \tag{36}$$

and b is a triple of constants.

When the value of β in terms of α and a from (33) and the value of γ from (35) are substituted in the differential equations for α in (32), these equations are readily integrated and have the solution

$$\frac{1}{AK}\alpha = \left(\frac{1}{2ABK^2} - \frac{B}{2A}W^2\right)a + \frac{B}{A}Wb + c, \tag{37}$$

where c is a fixed vector.

From equations (33), (35), (37), we find a, b, c as linear combinations of α, β, γ:

$$\begin{aligned} a &= -\frac{\partial K}{\partial s_1}\alpha - \frac{\partial K}{\partial s_2}\beta - \frac{\partial K}{\partial s_3}\gamma, \\ b &= -W\frac{\partial K}{\partial s_1}\alpha - W\frac{\partial K}{\partial s_2}\beta + \gamma, \\ c &= \left(\frac{B}{2A}W^2\frac{\partial K}{\partial s_1} + \frac{1}{2AK}\right)\alpha + \left(\frac{B}{2A}W^2\frac{\partial K}{\partial s_1} - \frac{1}{4BK}\right)\beta - \frac{B}{A}W\gamma, \end{aligned} \tag{38}$$

and hence obtain

$$\begin{aligned} &(a\,|\,a) = 0, \quad (b\,|\,b) = 1, \quad (c\,|\,c) = 0, \\ &(b\,|\,c) = 0, \quad (c\,|\,a) = B/A, \quad (a\,|\,b) = 0, \end{aligned} \qquad (a\,b\,c) = 1/2, \tag{39}$$

as the conditions on a, b, c equivalent to the initial conditions on α, β, γ. It should be noted, from (31), that B/A is a constant.

It is readily shown that equations (38) have the solutions

$$(40)\qquad \begin{aligned} (a \mid x) + a_0 &= -K, \\ (b \mid x) + b_0 &= -KW, \\ (c \mid x) + c_0 &= \frac{B}{2A}\left(KW^2 + \frac{1}{3B^2K}\right), \end{aligned}$$

defining x_1, x_2, x_3 in terms of B, K, W.

We proceed to find values of B, K in terms of U, V. For this purpose, it will be convenient to replace e^{-V} by $-\bar{V}^2$ in equations (8) and (9). These equations then become, after dropping the bars,

$$(41)\qquad \frac{\partial U}{\partial s_1} = -\frac{1}{V^2}, \qquad \frac{\partial U}{\partial s_2} = 0, \qquad \frac{\partial U}{\partial s_3} = 0,$$

$$(42)\qquad \frac{\partial V}{\partial s_1} = 0, \qquad \frac{\partial V}{\partial s_2} = -\tfrac{1}{2}AV, \qquad \frac{\partial V}{\partial s_3} = 0.$$

From (31), (34), (41), (42), it is readily verified that we may write

$$(43)\qquad B = -\frac{1}{UV^2}, \qquad K = \frac{V}{U}, \qquad UV \neq 0.$$

Denoting the left-hand sides of (40) by $a(x)$, $b(x)$, and $c(x)$, substituting the values of B and K from (43), and eliminating V, W, we obtain the equation

$$(44)\qquad [a(x)]^4U^6 + 3[b(x)]^2 + \frac{6A}{B}a(x)c(x) = 0,$$

which with (39) serves to define the required function U to within an additive and a multiplicative constant.

By means of the equations

$$a(x) = x_2' + ix_3', \qquad b(x) = \frac{2B}{A}ix_1', \qquad c(x) = \frac{B}{2A}(x_2' - ix_3'),$$

which by virtue of (39) represent a rigid motion, or, what is the same thing, by setting $a_0 = b_0 = c_0 = 0$, and taking as a, b, c the triples

$$(45)\qquad 0, 1, i, \qquad \frac{2B}{A}i, 0, 0, \qquad 0, \frac{B}{2A}, -\frac{B}{2A}i,$$

(44) reduces to the following normal form:

$$(x_2 + ix_3)^4 U^6 + 3(x_1^2 + x_2^2 + x_3^2) = 0. \tag{44a}$$

Thus here, too, the solution of the problem is essentially unique.

Equations (40), when B and K are replaced by their values from (43), represent the transformation from (x_1, x_2, x_3) to the curvilinear coordinates (U, V, W). From (41), (42), (36), it follows that the parametric surfaces for these coordinates form a triply orthogonal system and that the congruences of parametric curves consist precisely of the curves C_1, the curves C_2, and the curves C_3.

8. **The imaginary solution IIIa.** In this case we have

$$E = 0, \qquad F = -\tfrac{1}{2}A, \qquad A^2 + 4(B^2 + C^2) = 0, \qquad C \neq 0,$$

$$\begin{aligned}
&\frac{\partial A}{\partial s_1} = -AB, && \frac{\partial A}{\partial s_2} = A^2, && \frac{\partial A}{\partial s_3} = 0,\\
&\frac{\partial B}{\partial s_1} = \tfrac{1}{2}A^2 + B^2 - C^2, && \frac{\partial B}{\partial s_2} = AB, && \frac{\partial B}{\partial s_3} = 0,\\
&\frac{\partial C}{\partial s_1} = 2BC, && \frac{\partial C}{\partial s_2} = AC, && \frac{\partial C}{\partial s_3} = 0,
\end{aligned} \tag{46}$$

and equations (5) become

$$\begin{aligned}
&\frac{\partial \alpha}{\partial s_1} = A\beta, && \frac{\partial \alpha}{\partial s_2} = B\beta + C\gamma, && \frac{\partial \alpha}{\partial s_3} = C\beta - B\gamma,\\
&\frac{\partial \beta}{\partial s_1} = -A\alpha - C\gamma, && \frac{\partial \beta}{\partial s_2} = -B\alpha, && \frac{\partial \beta}{\partial s_3} = -C\alpha - \tfrac{1}{2}A\gamma,\\
&\frac{\partial \gamma}{\partial s_1} = C\beta, && \frac{\partial \gamma}{\partial s_2} = -C\alpha, && \frac{\partial \gamma}{\partial s_3} = B\alpha + \tfrac{1}{2}A\beta.
\end{aligned} \tag{47}$$

The principal normals of the curves C_3 are all parallel to one and the same isotropic plane, whose aspect is given by the fixed isotropic vector field

$$a = -\frac{A}{C^{1/2}}\alpha + \frac{2B}{C^{1/2}}\beta + 2C^{1/2}\gamma. \tag{48}$$

For the field of vectors

$$\xi = \frac{B}{C}\alpha + \frac{A}{2C}\beta$$

in the directions of the principal normals of the curves C_3 we find, in terms of the function L defined by the compatible equations

$$\frac{\partial L}{\partial s_1} = -\frac{A}{4C^{1/2}}, \qquad \frac{\partial L}{\partial s_2} = \frac{B}{2C^{1/2}}, \qquad \frac{\partial L}{\partial s_3} = \frac{1}{2}C^{1/2},$$

the relation

$$\xi = La + b,$$

where b is a fixed vector.

Using (48) and the equation which results from equating the two values of ξ, we may eliminate β, γ from the differential equations for α in (47). The resulting equations are found to have the integral

$$\frac{C^{1/2}}{A}\alpha = \left(-\frac{1}{2}L^2 - \frac{B}{AC^{1/2}}L + \frac{1}{8C}\right)a - \left(L + \frac{B}{AC^{1/2}}\right)b + c,$$

where c is a triple of constants.

We may now compute a, b, c in terms of α, β, γ:

$$\begin{aligned}
a &= 4\left(\frac{\partial L}{\partial s_1}\alpha + \frac{\partial L}{\partial s_2}\beta + \frac{\partial L}{\partial s_3}\gamma\right),\\
b &= \left(-4L\frac{\partial L}{\partial s_1} + \frac{B}{C}\right)\alpha + \left(-4L\frac{\partial L}{\partial s_2} + \frac{A}{2C}\right)\beta - 4L\frac{\partial L}{\partial s_3}\gamma,\\
c &= \left(-2L^2\frac{\partial L}{\partial s_1} + \frac{B}{C}L - \frac{A}{8C^{3/2}}\right)\alpha\\
&\quad + \left(-2L^2\frac{\partial L}{\partial s_2} + \frac{A}{2C}L + \frac{B}{4C^{3/2}}\right)\beta + \left(-2L^2\frac{\partial L}{\partial s_3} - \frac{1}{4C^{1/2}}\right)\gamma,
\end{aligned} \tag{49}$$

and hence obtain the conditions

$$\begin{aligned}
&(a\,|\,a) = 0, \quad (b\,|\,b) = -1, \quad (c\,|\,c) = 0,\\
&(b\,|\,c) = 0, \quad (c\,|\,a) = -1, \quad (a\,|\,b) = 0,
\end{aligned} \qquad (a\ b\ c) = -1, \tag{50}$$

which insure that α, β, γ are mutually perpendicular unit vector fields with the disposition of the axes.

Integration of equations (49) yields the relations

$$\begin{aligned}
(a\,|\,x) + a_0 &= 4L,\\
(b\,|\,x) + b_0 &= -2L^2 - \frac{1}{2C},\\
(c\,|\,x) + c_0 &= -\frac{2}{3}L^3 - \frac{1}{2C}L - \frac{B}{3AC^{3/2}},
\end{aligned} \tag{51}$$

where a_0, b_0, c_0 are arbitrary constants.

The directional derivatives of the function u defined by the equations

$$\frac{2B}{A} = u, \qquad \frac{2C}{A} = \epsilon(1+u^2)^{1/2}, \qquad (\epsilon = \pm i), \tag{52}$$

are $-6C^2/A, 0, 0$. Furthermore, it follows from (9) and (46) that $e^V = k(A^2C)^{1/3}$, where k is a constant, not zero. Comparison of the derivatives of U in (8) with those of u shows that U is a function of u. In particular, if we take $k=(2/\epsilon)^{1/3}$,

$$U = \frac{2}{3}\int \frac{du}{(1+u^2)^{5/6}}. \tag{53}$$

Denoting the left-hand sides of (51) by $a(x)$, $b(x)$, and $c(x)$, setting $B/A = u/2$, and eliminating L and C, we obtain the equation

$$\{[a(x)]^2 + 8b(x)\}^3u^2 + \{[a(x)]^3 + 12a(x)b(x) - 48c(x)\}^2 = 0, \tag{54}$$

which with (53) and the conditions (50) define the general solution U to within a multiplicative constant.

Without loss of generality we may take $a_0=b_0=c_0=0$ and as a, b, c the triples

$$0, 2, 2i, \qquad -i, 0, 0, \qquad 0, -\tfrac{1}{4}, \tfrac{1}{4}i. \tag{55}$$

Equation (54) then assumes the normal form

$$(z^2 - 2ix_1)^3u^2 + (z^3 - 3ix_1z + \tfrac{3}{2}\bar{z})^2 = 0, \tag{54a}$$

where $z=x_2+ix_3$ and $\bar{z}=x_2-ix_3$.

If in (53) we set $1+u^2=1/y^3$, we find that

$$U = -\int \frac{dy}{(1-y^3)^{1/2}}.$$

Thus, the function U involves an elliptic integral. The surfaces $U=$const. are, however, obviously algebraic.

C. The minimal surfaces

9. **The imaginary surfaces** IIa. We have to do here with equations (28) where $a_0=b_0=c_0=0$ and a, b, c have the values (30). When we set $U=u$, $1/C=v$, $B/C^2=w$, these equations become*

$$x_1 = i(2uvw + v), \quad z = vw, \quad \bar{z} = 4u^2vw + 4uv + \cot^{-1}(w/v), \quad vw \neq 0, \tag{56}$$

and represent a transformation from the coordinates (x_1, x_2, x_3), or $(x, z, \bar{z})$, to curvilinear coordinates (u, v, w). Corresponding to a generic point $(x, z, \bar{z})$

* Throughout the paper, $z=x_2+ix_3$ and $\bar{z}=x_2-ix_3$.

there are two sets of values of the curvilinear coordinates, namely, (u, v, w) and $(u+1/w, -v, -w)$. In other words, the transformation $u'=u+1/w$, $v'=-v$, $w'=-w$ has the effect of the identity.

From (56) we find, as the equations of the three families of parametric surfaces,

$$
\begin{aligned}
&S: \quad (ix_1 + 2uz)^2 = z \tan (4iux_1 + 4u^2z + \bar{z}),\\
(57) \qquad &S': \qquad\quad z = v^2 \cot ((x \mid x)/z + v^2/z),\\
&S'': \qquad\quad z = w^2 \tan ((x \mid x)/z + z/w^2).
\end{aligned}
$$

Since $u=U$, the surfaces S are actually the minimal surfaces $U=\text{const.}$, and inasmuch as $v=1/C$ and C is, according to §6, a function of V, the surfaces S' are actually the surfaces $V=\text{const.}$ Moreover, since

$$(58) \qquad (x_u \mid x_v) = 0, \qquad (x_u \mid x_w) = 2v^2, \qquad (x_v \mid x_w) = 0,$$

the surfaces S' cut the surfaces S'', as well as the surfaces S, orthogonally. The curves $u=\text{const.}$, $v=\text{const.}$ are, then, the curves C_3, and the curves $u=\text{const.}$, $w=\text{const.}$ are the curves C_2. The curves $v=\text{const.}$, $w=\text{const.}$, which, according to (57), are the parabolic circles in which the isotropic planes $z=\text{const.}$ intersect the spheres $(x \mid x)=\text{const.}$, are not, however, the curves C_1.

THEOREM 2. *The system of surfaces S, S', S'' admits the one-parameter group of rigid motions into itself for which the parabolic circles $v=\text{const.}$, $w=\text{const.}$ are the path curves. Each surface S', and each surface S'', admits this group of motions into itself, and the minimal surfaces S, which are permuted by it, are all congruent to the surface S_0: $u=0$, namely, $x_1^2+z \tan \bar{z}=0$.*

The surfaces S all contain the isotropic line L: $x_1=z=0$ and are tangent all along L to the isotropic plane Π: $z=0$. The group of motions in question consists of the "rotations" about L and may be represented by the equations

$$x_1' = x_1 + 2icz, \qquad z' = z, \qquad \bar{z}' = -4icx_1 + 4c^2z + \bar{z},$$

where c is the parameter. Since $(x' \mid x') = (x \mid x)$ and $z'=z$, the path curves of the group are actually the parabolic circles described.

The equations of the group in terms of the curvilinear coordinates, as found from (56), are, to within the identical transformation $u'=u+1/w$, $v'=-v$, $w'=-w$,

$$u' = u + c, \qquad v' = v, \qquad w' = w.$$

Hence, the theorem is proved.

From (56) follows the relation

$$ix_1 + 2uz + v = 0.$$

Hence, the curves C_3 are plane curves, and those on a specific surface S lie in parallel planes. In particular, the curves C_3 on the surface S_0 lie in the planes $x_1 = iv$ and have in these planes the equation $z = v^2 \cot \bar{z}$, and those on an arbitrary surface S are congruent to them.

Inasmuch as

$$iwx_1 + (1 + 2uw)z = 0,$$

the curves C_2 are plane curves lying in the Euclidean planes which pass through L. The general curve C_2 on S_0 is the intersection of the plane $iwx_1 + z = 0$ with the cylinder $z = w^2 \tan \bar{z}$ and the curves C_2 on the general surface S are congruent to those on S_0.

Using (58) and the formulas

$$(59) \qquad (x_u \mid x_u) = -4v^2w^2, \qquad (x_v \mid x_v) = (x_w \mid x_w) = -v^2/(v^2 + w^2),$$

we find that the curves C_1 are the curves $v = \text{const.}$, $2uv + \cot^{-1}(w/v) = \text{const.}$ Since, by (6), C is the torsion of these curves, it follows from (19) that they are twisted curves of constant torsion, and that this torsion is the same for all of them which lie on a given surface S'.

According to (59), the curves C_2 and C_3 on a surface S form an isometric system and v, w are isometric parameters.

10. **The imaginary surfaces** IIb. Consider the equations (40), where B and K are given by (43) in terms of U and V, a, b, c have the values (45), and $a_0 = b_0 = c_0 = 0$. According to (31), $B/2A = \pm i$. We may take $B/2A = i$, since, if $B/2A = -i$, changing the sign of W would yield the same results. If, then, U, V, W are replaced by $1/u$, v, w, the equations become

$$(60) \qquad x_1 = uvw, \qquad z = -uv, \qquad \bar{z} = uvw^2 + v^3/3u^3, \qquad uv \neq 0.$$

According to the last paragraph of §7, the parametric surfaces for the curvilinear coordinates (u, v, w) form a triply orthogonal system whose curves of intersection are the curves C_1, C_2, C_3. Corresponding to a generic point $(x, z, \bar{z})$ of space there are six sets of values of the curvilinear coordinates, of the form $(ku, v/k, w)$, where k takes on in turn the sixth roots of unity. Thus, the transformations $u' = ku$, $v' = v/k$, $w' = w$, $k^6 = 1$, have the same effect as the identity.

We find from (60), as the equations of the three families of parametric surfaces,

$$(61) \qquad \begin{aligned} S: &\quad z^4 + 3u^6(x \mid x) = 0, \\ S': &\quad 3z^2(x \mid x) + v^6 = 0, \\ S'': &\quad x_1 + wz = 0. \end{aligned}$$

The minimal surfaces S are algebraic surfaces of the fourth order. Each surface contains the isotropic line L: $z=x_1=0$ and is tangent all along L to the isotropic plane Π: $z=0$. The line at infinity in Π is a cuspidal edge of the surface with the plane at infinity as the cuspidal tangent plane; in particular, the point at infinity on L is a triple point at which the plane at infinity counts twice, and the plane Π once, as tangent planes. The origin O (on L) is a conical point with the isotropic cone at O as the tangent cone.

The surfaces S' are also surfaces of the fourth order. For each of them the line at infinity in Π is a cuspidal edge with the plane Π as the cuspidal tangent plane; in particular, the point at infinity on L is a triple point at which Π counts three times as tangent plane. There are no other singular points.

The surfaces S'' consist of the Euclidean planes through the line L.

THEOREM 3. *The triply orthogonal system of surfaces admits a two-parameter group of rigid motions into itself, consisting of the ∞^1 one-parameter groups of rotations about the lines of the pencil of lines which lies in the plane Π and has its vertex at O. Each surface of the system admits at least a one-parameter group of rigid motions into itself and each two surfaces of the same family are congruent.*

The one-parameter group of rotations about L has the equations

$$x_1' = x_1 - cz, \qquad z' = z, \qquad \bar{z}' = 2cx_1 - c^2z + \bar{z},$$

where c is the parameter. The corresponding equations in the curvilinear coordinates, to within one of the identical transformations, are

$$u' = u, \qquad v' = v, \qquad w' = w + c. \tag{62}$$

Hence each surface S, and each surface S', is carried into itself by the group of rotations about L. The path curves are the curves C_3 in which the surfaces S and S' intersect, namely, the parabolic circles in which the planes $z=$ const. cut the sphere $(x|x)=$ const.

The parabolic circles C_3 are the orthogonal trajectories of the planes S''. Since $w=-x_1/z$ is a harmonic function, these planes, too, constitute a family of harmonic minimal surfaces. The lines of flow are the curves C_3 and hence are the path curves of a one-parameter group of rigid motions.

One of the Euclidean lines through O in the plane Π is the x_1 axis. The one-parameter group of rotations about this axis has the equations $x_1'=x_1$, $z'=e^{\theta i}z$, $\bar{z}'=e^{-\theta i}\bar{z}$, or, in terms of the curvilinear coordinates, to within an identical transformation,

$$u' = e^{(2/3)\theta i}u, \qquad v' = e^{(1/3)\theta i}v, \qquad w' = e^{-\theta i}w. \tag{63}$$

The product of the general transformation (63) and the general trans-

formation (62) is the *two-parameter* group of rigid motions,

$$(64) \qquad u' = e^{(2/3)\theta i}u, \qquad v' = e^{(1/3)\theta i}v, \qquad w' = e^{-\theta i}w + c,$$

which we shall think of as consisting of the rotations (62) about L and the transformations

$$(65) \qquad u' = e^{(2/3)\theta i}u, \qquad v' = e^{(1/3)\theta i}v, \qquad w' - a = e^{-\theta i}(w - a).$$

For a fixed value of a these transformations are simply the rotations about the line through O with direction components $1, a, ai$. For, they leave O and the plane $w=a$, or $x_1+az=0$, fixed and hence leave fixed every point of the line in question.

Thus the two-parameter group (64) actually consists of the ∞^1 one-parameter groups of rotations about the ∞^1 lines passing through O and lying in the plane II. Obviously, every transformation of the group carries each family of surfaces of the triply orthogonal system into itself. Moreover, it is clear from (65) that, if two surfaces S, or two surfaces S', are given, there exists one rotation about each Euclidean axis which carries the one surface into the other. Thus, the theorem is completely established.

THEOREM 4. *The curves C_1 are plane quartic curves which lie in the planes S'' and are all congruent to the curve $z^3\bar{z}=1$ in the plane $x_1=0$. The curves C_2 are plane cubics which lie in the planes S'' and are all congruent to the curve $z^3=\bar{z}$ in the plane $x_1=0$. The curves C_3 are the parabolic circles in which the isotropic planes parallel to* II *cut the spheres with center at O.*

The curves C_1, which are the lines of flow in the physical problem, are the intersections of the planes S'' with the surfaces S' and hence are plane quartics. Since they are given by $v=\text{const.}$, $w=\text{const.}$, and since $v\neq 0$, there exists a unique rigid motion (64) which carries a given one of them into a second. They are, then, all congruent to the particular one $v=(-3)^{1/6}$, $w=0$, which lies in the plane $x_1=0$ and has the equation $z^3\bar{z}=1$.

Since each of the surfaces S contains the line L, the curves C_2 in which they are met by the planes S'' are plane cubics. These curves are given by $u=\text{const.}$, $w=\text{const.}$, where $u\neq 0$. It follows from (64) that they are all congruent to the particular one $u=(-1/3)^{1/6}$, $w=0$, which lies in the plane $x_1=0$ and has the equation $z^3=\bar{z}$.

The facts concerning the parabolic circles C_3 have already been established. Each two of them which lie on the same sphere are congruent, while two lying on different spheres are not.

It should perhaps be remarked that the plane cubics and the parabolic circles are the lines of curvature on the minimal surfaces S.

11. **The imaginary surfaces IIIa.** We are concerned in this case with equations (51), where $a_0=b_0=c_0=0$ and a, b, c have the values (55). When we set $2B/A=u$, $1/C=v^2$, $L=w/2$, the equations become

$$(66)\qquad x_1=-\tfrac{1}{2}i\,(v^2+w^2),\quad z=w,\quad \bar{z}=\tfrac{2}{3}uv^3+v^2w+\tfrac{1}{3}w^3,\quad 1+u^2\neq 0.$$

For a generic point of space there are two sets of values of the curvilinear coordinates thus introduced, namely, (u, v, w) and $(-u, -v, w)$. Thus, the transformation $u'=-u$, $v'=-v$, $w'=w$ is in effect the identity.

The equations of the three families of parametric surfaces are found to be

$$(67)\qquad \begin{aligned} S:&\quad u^2(2ix_1-z^2)^3=(z^3-3ix_1z+\tfrac{3}{2}\bar{z})^2,\\ S':&\quad 2ix_1-z^2=v^2,\\ S'':&\quad z=w. \end{aligned}$$

According to §8, u is a function of U. Hence, the surfaces S are actually the minimal surfaces $U=\text{const.}$ The surfaces S' are *not*, in this case, the survaces $V=\text{const.}$

The surfaces S' are parabolic cylinders which are tangent to the plane at infinity along the ideal line in the isotropic plane Π: $z=0$ and whose rulings are parallel to the isotropic line L: $z=x_1=0$. The surfaces S'' are the planes parallel to Π.

The w-curves are parabolic helices* lying on the parabolic cylinders S', except for the curve $v=0$:

$$(68)\qquad x_1=-\tfrac{1}{2}iw^2,\qquad z=w,\qquad \bar{z}=\tfrac{1}{3}w^3,$$

which is an isotropic cubic. The v-curves are plane cubics lying in the isotropic planes parallel to Π except for those for which $u=0$, which are Euclidean straight lines. Finally, the u-curves are the isotropic lines parallel to L.

The surfaces S for which $u\neq 0$ are algebraic surfaces of the sixth order. For each of them, the isotropic cubic K defined by (68) is a cuspidal edge, with the isotropic osculating planes of K as the cuspidal tangent planes, and the points of the line at infinity in the plane Π are all singular points, with the plane at infinity counting at least three times as tangent plane. The parametric curves on these surfaces are the parabolic circles and plane cubics just mentioned.

The surface $u=0$ is the cubic surface

$$z^3-3ix_1z+\tfrac{3}{2}\bar{z}=0,$$

counted twice. The cubic surface has the line at infinity in Π as a double

* A parabolic helix is a curve whose curvature and torsion are constant and in the ratio $\pm i$. The tangent indicatrix is a parabolic circle, lying in this case in a plane parallel to Π.

line, the two tangent planes at the ideal point in the direction a, 1, i being $iaz=1$ and the plane at infinity. Furthermore, it contains the isotropic cubic K as an asymptotic line. The remaining asymptotic lines are parabolic circles (w-curves) and straight lines (the special v-curves described above). Thus, this surface S is a *ruled* minimal surface, a parabolic helicoid.

The surfaces S for which $1+u^2=0$ have been excluded. They are the same surface, namely, the isotropic developable which is the tangent surface of the isotropic cubic K.

The differential coefficients of the general surface S, referred to v, w as parameters, where $D^2=EF-G^2$, are found to be

$$(69)\quad E=-v^2,\quad De=uv^2,\quad F=uv^2,\quad Df=v^2,\quad G=v^2,\quad Dg=-uv^2.$$

Since these coefficients are independent of w, the surface S admits the one-parameter group of rigid motions

$$(70)\qquad u'=u,\qquad v'=v,\qquad w'=w+c$$

into itself. In terms of the coordinates $(x, z, \bar{z})$, the equations of this group are

$$(71)\quad x_1'=x_1-icz-\tfrac{1}{2}ic^2,\quad z'=z+c,\quad \bar{z}'=2icx_1+c^2z+\bar{z}+\tfrac{1}{3}c^3.$$

Since each of these motions leaves fixed only the point at infinity in the direction of the isotropic line L and the tangent to the absolute at this point, the group may properly be described as the group of screw motions about the tangent to the absolute in question. Since $2ix_1-z^2$ is an absolute invariant, the path curves lie on the parabolic cylinders $v=$const. The path curves on the cylinders for which $v\neq 0$ are the parabolic helices (the w-curves), whereas those on the cylinder $v=0$ are isotropic cubics similar to the curve K.

THEOREM 5. *Each surface of the family of minimal surfaces is carried into itself by the group of screw motions about the tangent to the absolute at the ideal point of the isotropic line L. The path curves are the parabolic helices on the surface and the isotropic cubic K.*

From §8 and the relation $1/C=v^2$, it follows that u, v are functions of U, V. Hence, the w-curves (the parabolic helices and K) are actually the curves C_3.

The curves C_2 are the orthogonal trajectories of the w-curves on the surfaces S. As such, they are found to have the equation $uv+w=$const. Those on a specific surface S are all congruent, inasmuch as they cut the path curves of the rigid deformation of S into itself orthogonally. Hence, it suffices to consider the family of curves, one on each of the surfaces S, which is defined by the equation $w=-uv$. It is found that these curves are helices which lie

on cubic cylinders all congruent to the cylinder $\bar{z}=z^3$ and have on these cylinders varying pitches. When the curves are moved along the surfaces S by the group of screw motions, the point at infinity in the direction of the rulings of the cylinders on which they lie traces the axis of the group (the ideal line in the plane Π), since in its original position it is the ideal point in the direction 1, 0, 0, and hence a point on this axis. Finally, it may be shown that the curves C_2 are all tangent to the isotropic cubic K.*

On each surface S there is a single family of lines of curvature which covers the surface twice, and all of its members are tangent to the isotropic cubic K. For, the lines of curvature on an arbitrary surface S are the curves $r=\text{const.}$ and $s=\text{const.}$, where $w+iv=r$, $w-iv=s$, and since the identical transformation $u'=-u$, $v'=-v$, $w'=w$ interchanges r and s, the two families coincide. Furthermore, the locus $r=s$ is the curve $v=0$ or K, so that the tangency of the lines of curvature with K is indicated and readily verified.

Inasmuch as the lines of curvature on a specific surface S are all congruent, it suffices to consider those lines of curvature, one on each surface S, which are given by $r=0$ or $s=0$. It turns out that these lines of curvature are plane cubics all lying in the plane $x_1=0$ and all congruent to the cubic $\bar{z}=z^3$ in this plane. When they are moved along the surfaces S by the group of screw motions, their common plane envelopes the invariant parabolic cylinder $2ix_1-z^2=0$, since in its original position it is a tangent plane to this cylinder. All the lines of curvature on the surfaces S are, of course, congruent.

Employing the method of §9, we find as the parametric representation of the curves C_3, in terms of the parameter y of §8,

$$u=\frac{(1-y^3)^{1/2}}{y^{3/2}}, \qquad v=ky^{1/2}, \qquad w=\frac{k}{2}\int\frac{ydy}{(1-y^3)^{1/2}}+l,$$

where k and l are arbitrary constants. Thus, though the equipotential surfaces are algebraic, the lines of flow of the physical problem are transcendental.

D. Solution of the differential system

12. **First special case. Change of variables.** The crux of our problem, namely, the deduction of the solutions (I), (II), (III) of equations (12) and the proof that these are the only solutions, must finally be met. The three conditions of integrability on the derivatives of A given by (12a) yield two new relations, namely the finite relation

$$4ACE-4ABF+4B(B^2+C^2)=A^2B, \tag{72}$$

* In this discussion we have tacitly excluded the curves C_2 on the surface $u=0$. As we have already seen, these curves are straight lines.

and the differential equation

$$2\frac{\partial}{\partial s_3}(B^2+C^2) - A\left(\frac{\partial E}{\partial s_2}+\frac{\partial F}{\partial s_3}\right) = 4E(B^2+C^2). \tag{73}$$

First special case. Suppose that $B^2+C^2=0$. Differentiating with respect to s_1 and using (12b), we find that $ACE-ABF=0$, and therefore conclude, from (72), that $B=0$ and hence $C=0$. It follows from (12b) that $E=0$ and $F=0$. Thus, solution (Ib) is obtained.

Change of variables. We may assume henceforth that $B^2+C^2\neq 0$. Equations (12b) and (73) suggest the substitutions

$$2B = N^{1/2}\sin\phi, \quad 2C = N^{1/2}\cos\phi, \quad 4(B^2+C^2) = N. \tag{74}$$

Relation (72) becomes

$$4AE\cos\phi - 4AF\sin\phi + N\sin\phi = A^2\sin\phi,$$

and, when we adjoin the equation

$$4AE\sin\phi + 4AF\cos\phi - N\cos\phi = M,$$

the two relations yield values for E and F:

$$4AE = M\sin\phi + A^2\sin\phi\cos\phi, \quad 4AF = M\cos\phi - A^2\sin^2\phi + N. \tag{75}$$

Thus, we have introduced the new set of unknowns A, M, N, ϕ, in terms of which B, C, E, F are given by (74) and (75).

Equations (12a) become

$$\begin{aligned} 4\frac{\partial A}{\partial s_1} &= -2AN^{1/2}\sin\phi, \\ 4\frac{\partial A}{\partial s_2} &= A^2\sin^2\phi - M\cos\phi + 4A^2 + N, \\ 4\frac{\partial A}{\partial s_3} &= (M + A^2\cos\phi)\sin\phi, \end{aligned} \tag{76a}$$

and the equations resulting from (12b) are

$$\frac{\partial N}{\partial s_1} = A^2N^{1/2}\sin\phi, \qquad 2N^{1/2}\frac{\partial\phi}{\partial s_1} = -M - 2N\cos\phi, \tag{77a}$$

$$2N\frac{\partial\phi}{\partial s_2}+\frac{\partial N}{\partial s_3} = 4NE, \qquad 2N\frac{\partial\phi}{\partial s_3}-\frac{\partial N}{\partial s_2} = 4NF, \tag{77b}$$

where E and F are given by (75).

From the relations obtained by making the foregoing substitutions in the *first two* equations in (12c), we obtain the equations

$$2A\frac{\partial N}{\partial s_2} = M^2 + A^2M\cos\phi + 2N(A^2 - N), \tag{78}$$

$$2N^{1/2}\frac{\partial M}{\partial s_1} + 2A\frac{\partial N}{\partial s_3} = -A^2M\sin\phi. \tag{79}$$

It is clear from (77) and (78) that, in order to find finite values for all the derivatives of N and ϕ, we need the value of $\partial N/\partial s_3$. This derivative enters into the condition of integrability $(N; s_1, s_2)$, which is found to reduce to

$$(M - N\cos\phi)\frac{\partial N}{\partial s_3} + (A^2 - N)\frac{\partial N}{\partial s_2}\sin\phi + 4AN^2\sin\phi = 0. \tag{80}$$

Evidently, two cases arise, according as $M - N\cos\phi$ vanishes or not.

13. **Second special case.** If

$$M - N\cos\phi = 0, \tag{81}$$

it follows immediately from (78) and (80), since $N \neq 0$, that

$$\sin\phi(A^2 + N)[(A^2 - N)\cos^2\phi + 2(A^2 + N)] = 0.$$

(i) If $A^2+N=0$, we have from (75) and (81) that $E=0$ and $F=-A/2$. Substitution to these values in equations (12) leads immediately to solution (III).

(ii) If $\sin\phi=0$, the second equation in (77a), in conjunction with (81), yields $N=0$, a contradiction.

(iii) Suppose, finally, that

$$(A^2 - N)\cos^2\phi + 2(A^2 + N) = 0. \tag{82}$$

Differentiating this relation with respect to s_1, we find, since $N\sin\phi \neq 0$, that

$$(A^2 - 3N)\cos^2\phi = 0.$$

If $\cos\phi=0$, it follows from (82) that $A^2+N=0$ and we are thus led to solution (IIIb).

If $A^2-3N=0$, differentiation with respect to s_1 gives rise to an immediate contradiction.

14. **The derivatives in the general case.** We assume henceforth that $AN(M-N\cos\phi)\neq 0$, and introduce the quantity

$$P = \frac{S\sin\phi}{M - N\cos\phi}, \tag{83}$$

where

$$S = M^2 - 2Y^2 - A^4 \cos^2 \phi, \tag{84}$$

and

$$Y = A^2 + N = A^2 + 4(B^2 + C^2). \tag{85}$$

Since $M - N \cos \phi \neq 0$, equations (77), (78), and (80) may be solved for the derivatives of N and ϕ. From (79) follows then a finite value for $\partial M/\partial s_1$. Finally, when the substitutions (74) and (75) are carried out in (73) and the last equation in (12c), values for $\partial M/\partial s_2$ and $\partial M/\partial s_3$ result. Thus, the values of all twelve derivatives of the four unknowns A, M, N, ϕ are obtained, and from them may be computed the derivatives of Y and S.

The derivatives of A are already listed in (76a). Those of the other quantities are

$$\begin{aligned}
\frac{\partial N}{\partial s_1} &= A^2 N^{1/2} \sin \phi, \\
2A \frac{\partial N}{\partial s_2} &= M^2 + A^2 M \cos \phi + 2N(A^2 - N), \\
2A \frac{\partial N}{\partial s_3} &= -A^2(M + A^2 \cos \phi) \sin \phi + P,
\end{aligned} \tag{76b}$$

$$\begin{aligned}
2N^{1/2} \frac{\partial \phi}{\partial s_1} &= -M - 2N \cos \phi, \\
4AN \frac{\partial \phi}{\partial s_2} &= (A^2 + 2N)(M + A^2 \cos \phi) \sin \phi - P, \\
4AN \frac{\partial \phi}{\partial s_3} &= M^2 + (A^2 + 2N) M \cos \phi + 2A^2 N \cos^2 \phi,
\end{aligned} \tag{76c}$$

$$\begin{aligned}
2N^{1/2} \frac{\partial M}{\partial s_1} &= A^4 \sin \phi \cos \phi - P, \\
4AN \frac{\partial M}{\partial s_2} &= (3A^2 + 10N + A^2 \cos^2 \phi) A^2 M \\
&\quad + (5A^2 + 4N + A^2 \cos^2 \phi) A^4 \cos \phi \\
&\quad + [M + 2(A^2 - N) \cos \phi] S + \sin \phi (2N - A^2) P, \\
4AN \frac{\partial M}{\partial s_3} &= -A^4 (M + A^2 \cos \phi) \sin \phi \cos \phi + (3M + A^2 \cos \phi) P,
\end{aligned} \tag{76d}$$

$$\text{(76e)} \qquad \frac{\partial Y}{\partial s_1} = 0, \qquad 2A\frac{\partial Y}{\partial s_2} = S + 7A^2Y, \qquad 2A\frac{\partial Y}{\partial s_3} = P,$$

$$N^{1/2}\frac{\partial S}{\partial s_1} = -MP,$$

$$\text{(76f)} \quad 2AN\frac{\partial S}{\partial s_2} = [6A^2(M + A^2\cos\phi)\cos\phi + (6A^2 - 8N)Y]A^2Y + kS + lP,$$

$$2AN\frac{\partial S}{\partial s_3} = (3M^2 + A^2M\cos\phi - 4NY)P,$$

where k and l are functions of A, N, M, ϕ whose explicit values we shall not need.

We distinguish two cases, according as P is, or is not, zero.

15. **Third special case.** If $P=0$, it follows from (83) that $S\sin\phi=0$.

(i) If $\sin\phi=0$, then $B=0$ and since $N\neq 0$, $C\neq 0$. From (75), $E=0$, and from (12b), $AF+C^2=0$. We are thus led to solution (Ia).

(ii) If $\sin\phi\neq 0$ and $S=0$, it follows from the equation $\partial S/\partial s_2=0$ that $YZ=0$, where Z may be read off from (76f).

Suppose that $Y=0$. From $S=0$, it follows, since $M-N\cos\phi\neq 0$, that $M=A^2\cos\phi$. Hence, it is found from (75) and (74) that $AE=-2BC$ and $AF=2B^2$, and we arrive at solution (II).

If $Y\neq 0$, then

$$Z \equiv 3A^2(M + A^2\cos\phi)\cos\phi + (3A^2 - 4N)Y = 0.$$

Eliminating M from this equation and the equation $S=0$, we obtain the relation

$$T \equiv (3A^2 - 4N)^2Y - 42A^4N\cos^2\phi = 0.$$

By means of $S=0$, $Z=0$, $T=0$, it follows from the equation $\partial T/\partial s_2=0$ that $3A^2-4N=0$ or $3A^2+10N=0$. But differentiation of either of these equations with respect to s_1 leads to the contradiction $\sin\phi=0$. Thus, the discussion of this special case is complete.

16. **General case. Conclusion.** When $P\neq 0$, then $Y\neq 0$, for if Y were zero, it would follow from (76e) that $P=0$. Consequently we have in the present case $ANSY(M-N\cos\phi)\sin\phi\neq 0$.

The integrability conditions on A we know to be satisfied. Those on N may be replaced by the conditions on Y. The conditions $(Y; s_1, s_2)$ and $(\phi; s_1, s_3)$ are found to be identities, and the conditions $(Y; s_1, s_3)$ and $(\phi; s_1, s_2)$ both reduce to the equation

$$\frac{\partial P}{\partial s_1} = 0, \tag{86a}$$

which, when P is replaced by its value from (83), can be written in the form

$$P \sin\phi = M^2 \cos\phi + (N - 2A^2 \sin^2\phi)M + A^4 \sin^2\phi \cos\phi - 2N^2 \cos\phi. \tag{86b}$$

The condition of integrability $(\phi; s_2, s_3)$ reduces, by (86b), to

$$AN \frac{\partial \log P}{\partial s_3} \sin\phi = M^2 \cos\phi + (N - A^2 \sin^2\phi)M - 2N^2 \cos\phi.$$

On the other hand, we find that the value of $\partial \log P/\partial s_3$, computed from (83), becomes, by virtue of (86b),

$$A(M - N\cos\phi) \frac{\partial \log P}{\partial s_3} = (M^2 - 2NY) \sin\phi.$$

Eliminating $\partial \log P/\partial s_3$, we obtain the relation

$$M^3 \cos\phi - A^2 \sin^2\phi\, M^2 + (A^2 N \sin^2\phi - 3N^2)M \cos\phi + 2A^2N^2 \sin^2\phi + 2N^3 = 0. \tag{87}$$

If we differentiate (86b) with respect to s_1, remembering that $\partial P/\partial s_1 = 0$, and eliminate P from the resulting equation by means of (86b), we get

$$M^3 \cos 2\phi - 8A^2 \sin^2\phi \cos\phi\, M^2 - (3N^2 \cos 2\phi + A^2 r)M + (2N^3 + A^2 s)\cos\phi = 0, \tag{88}$$

where r and s are polynomials in A, N, $\cos^2\phi$.

When the value of P from (83) is substituted in (86b) and the resulting equation is cleared of fractions, there is obtained an equation of the form $H = 0$, where H is a cubic polynomial in M whose leading term is $M^3 \cos\phi$. Subtraction of this equation from (87) yields the simple relation

$$M^2 - YM \cos\phi - 2NY = 0. \tag{89}$$

LEMMA 1. *The condition on A and N that the three equations* (87), (88), (89), *considered as equations in M and ϕ, be compatible is not an identity in A and N.*

LEMMA 2. *There exists no functional relation between A and N.*

Once these contradictory lemmas are proved it will follow that no further solution of the differential system exists.

To prove Lemma 2, we assume that A and N are functionally related: $\Phi(A, N) = 0$. Inasmuch as $\partial\Phi/\partial s_1 = 0$ and $\partial Y/\partial s_1 = 0$, it follows, since $\partial A/\partial s_1 \neq 0$, that $\Phi \equiv f(Y)$. But, if we differentiate the relation $f(Y) = 0$ with respect to s_3, we get $P = 0$, a contradiction.

To prove that the condition on A and N described in Lemma 1 is not an identity in A and N, it suffices to show that it is not an identity in N when $A=0$. In this case, the expressions on the left-hand sides of (87), (88), and (89) become

$$a \equiv M^3 \cos\phi - 3N^2M \cos\phi + 2N^3,$$
$$b \equiv M^3 \cos 2\phi - 3N^2M \cos 2\phi + 2N^3 \cos\phi,$$
$$c \equiv M^2 - NM \cos\phi - 2N^2,$$

and satisfy the relations

$$a \cos 2\phi - b \cos\phi \equiv -2N^3 \sin^2\phi, \tag{90a}$$

$$a \equiv (M \cos\phi + N \cos^2\phi)c + N^2 d, \tag{90b}$$

where

$$d \equiv M(\cos^2\phi - 1)\cos\phi + 2N(\cos^2\phi + 1).$$

From (90a) it follows that, if the equations $a=0$, $b=0$ have a common solution in M and ϕ, the solution must be $M=M$, $\sin\phi=0$. If this solution is also to satisfy the equation $c=0$, it must, by (90b), satisfy $d=0$. But, when $\sin\phi=0$, then $d\equiv 4N$, and the condition $N=0$ is not an identity in N. Thus, Lemma 1 is established.

E. Special cases

17. **The special case $\Delta_1 V=0$.** When $\Delta_1 V=0$ and V is not a constant, ∇V is an isotropic vector field which, since $\Delta_1(U, V)=0$, is "orthogonal" to α. There then exists a second isotropic vector function η such that the three vector fields

$$\alpha = \frac{\nabla U}{(\Delta_1 U)^{1/2}}, \qquad \xi = \nabla V, \qquad \eta \tag{91}$$

enjoy the relations

$$\begin{aligned} &(\alpha \mid \alpha) = 1, \quad (\xi \mid \xi) = 0, \quad (\eta \mid \eta) = 0, \\ &(\alpha \mid \xi) = 0, \quad (\alpha \mid \eta) = 0, \quad (\xi \mid \eta) = -i, \end{aligned} \qquad (\alpha\, \xi\, \eta) = 1, \tag{92}$$

provided merely that, to insure $+1$ as the value of $(\alpha\, \xi\, \eta)$, we are permitted, if necessary, to change the sign of U.

The nonholonomic derivatives corresponding to the three vector fields α, η, ξ are

$$\frac{\partial f}{\partial t_1} = \sum \alpha_i \frac{\partial f}{\partial x_i}, \qquad \frac{\partial f}{\partial t_2} = \sum \eta_i \frac{\partial f}{\partial x_i}, \qquad \frac{\partial f}{\partial t_3} = \sum \xi_i \frac{\partial f}{\partial x_i}. \tag{93}$$

In terms of them, the gradient of the function f has the form

$$\nabla f = \alpha \frac{\partial f}{\partial t_1} + i\xi \frac{\partial f}{\partial t_2} + i\eta \frac{\partial f}{\partial t_3}, \tag{94}$$

as is readily shown by means of the relations

$$\widetilde{\xi\eta} = \alpha, \qquad \widetilde{\alpha\xi} = i\xi, \qquad \widetilde{\eta\alpha} = i\eta.$$

It follows from (91) and (93) that

$$\begin{aligned} &\frac{\partial U}{\partial t_1} = e^V, \qquad \frac{\partial U}{\partial t_2} = 0, \qquad \frac{\partial U}{\partial t_3} = 0, \\ &\frac{\partial V}{\partial t_1} = 0, \qquad \frac{\partial V}{\partial t_2} = -i, \qquad \frac{\partial V}{\partial t_3} = 0. \end{aligned} \tag{95}$$

The equations of variations of α, ξ, η, in terms of the nonholonomic derivatives, have, in view of relations (92), the general form

$$\frac{\partial \alpha}{\partial t_j} = A_j\xi + B_j\eta, \qquad \frac{\partial \xi}{\partial t_j} = iB_j\alpha + C_j\xi, \qquad \frac{\partial \eta}{\partial t_j} = iA_j\alpha - C_j\eta, \tag{96}$$

where $j = 1, 2, 3$.

From these equations are obtained the conditions of integrability for the nonholonomic derivatives. When these conditions are applied to the derivatives of U and V given in (95), the following relations result:

$$\begin{aligned} &B_3 = 0, \qquad C_3 = 0, \qquad B_2 + C_1 = 0, \\ &A_1 = 1, \qquad B_1 = 0, \qquad A_3 - B_2 = 0. \end{aligned}$$

To these is to be added, in accordance with (1), the condition $A_3 + B_2 = 0$ which guarantees that $\Delta_2 U = 0$. Consequently, all of the functions A_j, B_j, C_j vanish except A_2 and C_2, which we shall henceforth denote by A and C.

Equations (96) now become

$$\begin{aligned} &\frac{\partial \alpha}{\partial t_1} = \xi, \qquad \frac{\partial \alpha}{\partial t_2} = A\xi, \qquad \frac{\partial \alpha}{\partial t_3} = 0, \\ &\frac{\partial \xi}{\partial t_1} = 0, \qquad \frac{\partial \xi}{\partial t_2} = C\xi, \qquad \frac{\partial \xi}{\partial t_3} = 0, \\ &\frac{\partial \eta}{\partial t_1} = i\alpha, \qquad \frac{\partial \eta}{\partial t_2} = iA\alpha - C\eta, \qquad \frac{\partial \eta}{\partial t_3} = 0, \end{aligned} \tag{97}$$

and the conditions of integrability reduce to

$$
\begin{aligned}
&\frac{\partial}{\partial t_2}\frac{\partial f}{\partial t_1} - \frac{\partial}{\partial t_1}\frac{\partial f}{\partial t_2} = -i\frac{\partial f}{\partial t_1} + A\frac{\partial f}{\partial t_3},\\
&\frac{\partial}{\partial t_3}\frac{\partial f}{\partial t_2} - \frac{\partial}{\partial t_2}\frac{\partial f}{\partial t_3} = -C\frac{\partial f}{\partial t_3},\\
&\frac{\partial}{\partial t_1}\frac{\partial f}{\partial t_3} - \frac{\partial}{\partial t_3}\frac{\partial f}{\partial t_1} = 0.
\end{aligned}
\tag{98}
$$

Equations (97) are completely integrable if A and C satisfy the differential equations

$$
\frac{\partial C}{\partial t_1} = 0, \qquad \frac{\partial C}{\partial t_3} = 0, \qquad \frac{\partial A}{\partial t_1} = C + i, \qquad \frac{\partial A}{\partial t_3} = 0,
\tag{99}
$$

and this system of equations, unlike the corresponding system in the general case, is readily integrated. In fact, C is an arbitrary function of V which we choose to write in the form $C = iF'(V)$, where the prime denotes differentiation, and A has, then, the value $A = i(1+F')e^{-V}U + e^{F}\Phi'$, where $\Phi(V)$ is a second arbitrary function of V.

It follows from (97) that

$$
\alpha - (e^{-V}U + ie^{F}\Phi)\xi = a, \qquad e^{F}\xi = b,
\tag{100}
$$

where a and b are constant vector fields, which, on account of (92), must satisfy the relations

$$
(a \mid a) = 1, \qquad (b \mid b) = 0, \qquad (a \mid b) = 0.
\tag{101}
$$

The nonholonomic derivatives of the functions whose gradients are the expressions on the left-hand sides of equations (100) are obtainable by comparison of these expressions with (94). Thus, the integrals of the equations are found to be

$$
e^{-V}U - i\int e^{F}\Phi dV = (a \mid x) + a_0, \qquad \int e^{F}dV = (b \mid x) + b_0.
$$

Denoting $(a|x)+a_0$, $(b|x)+b_0$ by $a(x)$, $b(x)$, and eliminating V, we obtain the equation

$$
U = f(b(x))a(x) + \phi(b(x)),
\tag{102}
$$

where, since $F(V)$ and $\Phi(V)$ are arbitrary functions, $f(y)$ and $\phi(y)$ are arbitrary functions, the first of which is not a constant. This equation, subject to the attendant conditions (101), defines the general solution U in this special case. For, it is readily found that

(103) $\nabla U = fa + [f'a(x) + \phi']b, \quad \Delta_2 U = 0, \quad V = \log f, \quad \nabla V = (f'/f)b,$

whence the conclusion follows.

It is clear from (101) that there exists a rigid motion which transforms $-a(x)$ and $2b(x)$ respectively into x_1 and x_2+ix_3. By means of this rigid motion or, what is the same thing, by setting a_0 and b_0 equal to zero and taking a and b as the triples $-1, 0, 0$ and $0, 1/2^{1/2}, i/2^{1/2}$, we reduce equation (102) to the normal form

(102a) $$U = f_1(z)x_1 + \phi_1(z),$$

where f_1 and ϕ_1 are arbitrary functions, the first of which is not a constant.

The surfaces $U=$const. defined by (102a) are cylinders with isotropic rulings parallel to the line $L: x_1=0, z=0$, except in the case $f_1(z)=1/z, \phi_1(z)=0$, when they consist of the Euclidean planes through L. The surfaces $V=$const. are always the isotropic planes $z=$const. parallel to L, and the lines of flow—the orthogonal trajectories of the surfaces $U=$const.—are parabolic circles lying in these planes and cut by them from spheres whose centers are in the plane $\bar{z}=0$. The group of rotations about L, or that about any line parallel to L, carries the totality of surfaces of all the families $U=$const. into itself.

We have thus far excluded the case in which V is a constant, and consequently have demanded that $f_1(z)$ be not a constant. As a matter of fact, the function

(104) $$U = x_1 + g(z),$$

where $g(z)$ is arbitrary, is a solution of our problem, as is clear from (103). Moreover, it gives a normal form for all solutions in the case $V\equiv$const. For, since in this case $\Delta_1 U$ is a constant, not zero, the surfaces $U=$const. are geodesically parallel and, inasmuch as they are minimal, they must be Euclidean planes or non-isotropic minimal developables. A minimal developable has isotropic rulings and hence, if it has Euclidean tangent planes, must be a cylinder with isotropic rulings. Therefore, in any case, the parallel surfaces are given by $U=$const., where U is of the form (102a) subject to the condition $f_1(z)\neq 0$. But, since $\Delta_1 U$ must be constant, it follows from (103) that (102a) reduces to (104) and the proof is complete.

Hence, in case $\Delta_1 U$ is constant, not zero, the minimal surfaces $U=$const. consist of parallel planes or special parallel cylinders with isotropic rulings. There are no surfaces $V=$const., and the lines of flow in both cases are straight lines.

HARVARD UNIVERSITY,
CAMBRIDGE, MASS.

THE GEOMETRY OF FIELDS OF LINEAL ELEMENTS*

BY

JOHN DE CICCO

1. **Introduction.** We shall begin by considering certain simple operations or transformations on the oriented lineal elements of the plane. A *turn* T_α converts each element into one having the same point and making a fixed angle α with the original direction. By a *slide* S_k, the line of the element remains the same and the point moves along the line a fixed distance k. These transformations together form a continuous group of three-parameters which we call the *whirl group* W_3. The group of whirls is isomorphic to the group of rigid motions M_3. These two three-parameter groups are commutative and together generate a continuous group of six-parameters which we term the *whirl-motion group* G_6. In preceding papers (see the bibliography at the end of this paper), Kasner and the author developed the geometry of this group G_6. In this paper, which is a continuation of the paper by the author *The differential geometry of series of lineal elements*, these Transactions, vol. 46 (1939), pp. 348–361, we shall give the differential geometry of fields of lineal elements with respect to the whirl-motion group G_6.

A set of ∞^1 elements is called a *series*; this includes a union (curve or point) as a special case. A collection of ∞^2 elements is termed a *field*, which of course corresponds to a differential equation of the first order, $F(x, y, y') = 0$. The totality of ∞^3 elements of the plane is called the *opulence* (as defined by Kasner).

In the earlier paper, we considered the tangent turbines, the osculating flat fields, and the osculating limaçon (circular) series of a given series S. We defined the *curvature* κ and the *torsion* τ of any series. The curvature $\bar{\kappa}$ and the torsion τ of a series $\bar{S}$ conjugate to a given series S are given by the formulas $\bar{\kappa} = \kappa/\tau$, $\bar{\tau} = 1/\tau$. We proved the fundamental result that *any two general (equiparallel) series which have their curvatures and torsions the same functions of the angle u (arc length s) are equivalent under the whirl-motion group* G_6. This result establishes the intrinsic equations of any series in the geometry of the whirl-motion group G_6.

In the present paper, we shall derive the analogues of some of the classic theorems for a surface in a euclidean three-dimensional space. In particular,

* Presented to the Society, March 26, 1937, under the title *The differential geometry of series of lineal elements*, and September 6, 1938, under the title *Asymptotic directions of a field of lineal elements*; received by the editors November 16, 1939.

we shall consider the Meusnier, the Euler, the Joachimsthal, and the Beltrami-Enneper theorems for a field in the geometry of the whirl-motion group G_6. The theory of geodesic series (minimum curvature) will be developed. We shall define the *gaussian curvature* K of a field. Finally the theory of conjugate fields will be considered.

2. **The tangent turbines of a series.** A series which consists of ∞^1 nonparallel (parallel) elements is called a *general* (*equiparallel*) *series*. A general series is given by $v=v(u)$, $w=w(u)$, whereas an equiparallel series is given by $u=c$, $w=w(v)$, where c is a constant. A general series possesses a point-union and a line-union, whereas an equiparallel series possesses only a point-union.

A *turbine* is the series which is obtained by applying a turn T_α to the elements of an oriented circle (the outer circle). It is *nonlinear* or *linear* according as this base circle is not or is a straight line.

A nonlinear turbine is a general series. Its point-union is a circle (the outer circle), and its line-union is also a circle (the inner circle). These two circles are concentric, and their common center is called the *center* of the turbine.

From the preceding remarks, we find that a nonlinear turbine may be constructed by applying a slide S_s to the elements of an oriented circle (the inner circle). Thus the equations of a nonlinear turbine are

$$\begin{aligned} v &= a\cos u + b\sin u + r, \\ w &= -a\sin u + b\cos u + s, \end{aligned} \tag{1}$$

where (a, b) are the cartesian coordinates of the center, r is the radius of the inner circle, and s is the constant distance of the slide S_s. We call $T(a, b, r, s)$ a set of nonlinear turbine coordinates.

A linear turbine is an equiparallel series whose base curve is a straight line. The equations of a linear turbine are

$$u = U - \omega, \qquad v\cos\omega + w\sin\omega = V, \tag{2}$$

where (U, V) are the hessian coordinates of the base line and ω is the constant angle of the turn T_ω. We call $T(U, V, \omega)$ a set of linear turbine coordinates. Obviously

$$T(U, V, \omega) = T(U+\pi, -V, \omega+\pi).$$

The *angle* u_2-u_1 between any two elements is the angle between their lines. Two elements are *parallel* or *supplementary* (*antiparallel*) according as the angle between them is 0 or π. The *distance* $[(v_2-v_1)^2+(w_2-w_1)^2]^{1/2}$ between two *parallel* elements is the distance between their points.

Two parallel elements are on a unique linear turbine. Two nonparallel elements are contained in a unique nonlinear turbine T. The center of T is the intersection between the perpendicular bisector of the segment determined by the points of E_1 and E_2, and the angle bisector of the angle determined by the oriented lines of E_1 and E_2.

Two series S_1 and S_2 are said to be *tangent* (or to have contact of the first order) at a common element E if they have two (but not three) consecutive elements in common at E. The two series S_1 and S_2 are said to be *osculating* (or to have contact of the second order) at E if they have three (but not four) consecutive elements in common at E.

If a one-parameter family of series has the property that consecutive series have a common element, the family is called a *set of enveloping series*. The locus of intersection of consecutive series is termed the *envelope*. It is easy to prove that *any series S_t of a set of enveloping series is tangent to the envelope S at any one of their common elements*.

In the remainder of the paper, an accent will always mean total differentiation with respect to u unless otherwise specified.

THEOREM 1. *The tangent turbines of a general series are the ∞^1 nonlinear turbines whose parameter values are*

$$(3)\qquad \begin{aligned} a &= -v' \sin u - w' \cos u, & b &= v' \cos u - w' \sin u, \\ r &= v + w', & s &= -v' + w. \end{aligned}$$

On the other hand, the tangent turbines of an equiparallel series S are the ∞^1 linear turbines all of which possess the common direction of S and whose base lines are tangent to the base curve of S.

It may be observed that two series S_1 and S_2 are tangent at a common element E if and only if they have the same tangent turbine at E.

3. **Conjugate series.** Two turbines T and $\overline{T}$ are said to be *conjugate* if they have the same circle as point-locus and the elements of the two turbines are symmetrically related to the elements of the circle. Two series S and $\overline{S}$ are said to be *conjugate* if there exists a one-to-one correspondence between their elements in such a way that the tangent turbines of the two series at the corresponding elements are conjugate turbines.

The conjugate turbines $\overline{T}_1$ and $\overline{T}_2$ of two given turbines T_1 and T_2 (not both linear) do or do not possess a common element according as T_1 and T_2 do or do not possess a common element. The conjugate turbines of two intersecting linear turbines never possess a common element.

THEOREM 2. *For any general series S, there always exists one and only one conjugate series $\overline{S}$ which either consists of one element or is a general series. This*

series $\overline{S}$ is given by the equations

$$(4)\qquad \cos\bar{u} = \frac{-a'r' + b's'}{a'^2 + b'^2}, \qquad \sin\bar{u} = \frac{-a's' - b'r'}{a'^2 + b'^2},$$
$$\bar{v} = a\cos\bar{u} + b\sin\bar{u} + r, \qquad \bar{w} = -a\sin\bar{u} + b\cos\bar{u} - s,$$

where (a, b, r, s) are the parameter values of the tangent turbines of S.

If an equiparallel series S is not a turbine, then there is no series conjugate to it.

4. **The osculating flat fields of a series.** A *nonlinear flat field* consists of the ∞^2 elements cocircular with a given element, called the *central element.* The equation of a nonlinear flat field Π is

$$(5)\qquad w = (v - \bar{v})\cot(u - \bar{u})/2 - \bar{w},$$

where $(\bar{u}, \bar{v}, \bar{w})$ are the hessian coordinates of the central element $\overline{G}$ of Π. We call $\Pi(\bar{u}, \bar{v}, \bar{w})$ a set of nonlinear flat field coordinates.

A *linear flat field* is the set of ∞^2 elements on ∞^1 parallel straight lines. Any linear field is given by $u = \text{const}$.

The invariants between two nonlinear flat fields are identical with those between their central elements.

In a given flat field, there are ∞^2 turbines. The turbines which are contained in a nonlinear flat field Π are those whose conjugate turbines possess the central element $\overline{G}$ of Π. The turbines which are contained in a linear flat field Π are the linear turbines which have the common direction of Π.

Three parallel elements determine a unique linear flat field. Three elements which are not all parallel and which do not lie on one turbine determine a unique nonlinear flat field Π. The central element $\overline{G}$ of Π is the single intersection of the conjugate turbines of the three turbines which pass through these elements.

Two elements of a flat field Π determine a turbine which lies entirely in Π. Two flat fields (not both linear) intersect in a turbine. Two linear flat fields have no common elements.

The flat field which has three consecutive elements in common with a series S at an element E of S is called the *osculating flat field* of S at E.

THEOREM 3. *The osculating flat fields of a general series S are the nonlinear flat fields whose central elements are the elements of the series $\overline{S}$ conjugate to S.*

If $\overline{S}$ consists of only one element $\overline{G}$, then S is contained in the nonlinear flat field whose central element is $\overline{G}$. In this case, we shall say that S is a *coflat series.*

Any equiparallel series S has one and only one osculating flat field, namely, the linear flat field in which it is contained.

5. **The osculating limaçon series of a general series.** Let T be a nonlinear turbine, let $\overline{G}$ be a fixed element on the conjugate turbine $\overline{T}$ of T, and let γ be a real number. Let O be the point of $\overline{G}$, and let P be the point of any element E of T. On the line (OP), let us select the points P_i $(i=1, 2)$ such that the distance $d(P, P_i)=2\gamma$. Let E_i be the element whose point is P_i and whose direction is that of E. By this construction, to each element E of T there are associated two parallel elements E_1 and E_2. The totality of these elements E_1, E_2 is called a *limaçon series* with *central turbine* T and *radius* γ.

Upon letting C and D denote

$$C = -2\gamma \sin \bar{u}/2, \qquad D = 2\gamma \cos \bar{u}/2, \tag{6}$$

we find that the equations of a limaçon series are

$$\begin{aligned} v &= A \cos u + B \sin u + C \cos u/2 + D \sin u/2 + R, \\ w &= -A \sin u + B \cos u - C \sin u/2 + D \cos u/2 + S, \end{aligned} \tag{7}$$

where (A, B, R, S) are the parameters of the central turbine T, $\bar{u}$ is the normal angle of the fixed element $\overline{G}$, and γ is the radius of the limaçon series. We call $L(A, B, C, D, R, S)$ a set of limaçon series coordinates. Obviously

$$L(A, B, C, D, R, S) = L(A, B, -C, -D, R, S).$$

A limaçon series L is contained in the nonlinear flat field Π whose central element is $\overline{G}$. The centers of the tangent turbines of L, which of course are in Π, are on a circle with center (A, B) and radius γ. We call this the *associated circle* of L. A limaçon series is uniquely determined by its flat field and its associated circle.

Three elements no two of which are parallel and which do not all lie on one turbine determine *four* limaçon series. Three elements only two of which are parallel determine *two* limaçon series. The flat field Π of these limaçon series is the one determined by the given elements. Their associated circles are those which are tangent to all three of the angle bisectors of the angles formed by each of the oriented lines of these elements with the oriented line of the central element $\overline{G}$ of Π. In the first case, there are four circles, namely, the inscribed and escribed circles of the triangle formed by these lines. In the second case, there are only two circles since two of these three lines are parallel.

THEOREM 4. *The osculating limaçon series of a general series S are those whose parameter values are*

$$
\begin{aligned}
&A = a + 2r' \sin u + 2s' \cos u, && B = b - 2r' \cos u + 2s' \sin u,\\
(8)\quad &C = -4r' \sin u/2 - 4s' \cos u/2, && D = 4r' \cos u/2 - 4s' \sin u/2,\\
&R = r + 2s', && S = s - 2r',
\end{aligned}
$$

where (a, b, r, s) *are the parameters of the tangent turbines of* S.

The envelope of the central turbines of the osculating limaçon series of a general series S is called the *series of curvature* of S. This series is given by

$$U = u + \pi, \qquad V = v + 2w', \qquad W = -2v' + w. \tag{9}$$

THEOREM 5. *The tangent turbines and the central turbines of any general series* S *have in common the series of curvature of* S.

6. **The osculating circular series of an equiparallel series.** An equiparallel series whose point-union is a circle with center (A, B) and radius γ is called a *circular series* with *center* (A, B) and *radius* γ.

The osculating circular series of an equiparallel series S are those which possess the common direction of S and whose circles are the osculating circles of the point-union of S.

7. **The curvature and torsion of a general series.** The *curvature* κ at an element E of a general series S is defined by the formula

$$\kappa = (r'^2 + s'^2)^{1/2}, \tag{10}$$

where (a, b, r, s) are the parameters of the tangent turbine of S at E.

The quantity κ is one half of the radius of the osculating limaçon series L of S at E, and also it is one half of the distance between the centers of the tangent and central turbines of S at E.

The *torsion* τ at an element E of a general series S is defined by the formula

$$\tau = d\bar{u}/du, \tag{11}$$

where u and $\bar{u}$ are the normal angles of the element E of S and the element $\overline{E}$ which is the central element of the osculating flat field of S at E.

The torsion τ at an element E of a general series S is the rate of change of the angle of the osculating flat field per unit radian measure of the angle of the element E.

8. **The curvature of an equiparallel series.** The *curvature* $\kappa = 1/\gamma$ at an element E of an equiparallel series S is defined by the formula

$$\kappa = \frac{1}{\gamma} = \frac{w''}{(1 + w'^2)^{3/2}}, \tag{12}$$

where the accent denotes differentiation with respect to v.

The quantity $\gamma = 1/\kappa$ is the radius of the osculating circular series of S at E.

The *torsion* τ of an equiparallel series is taken to be zero.

9. **The osculating spherical fields of a general series.** Let E denote an element, and let Π denote a nonlinear flat field with central element $\overline{G}$. On the oriented line of E, construct the element G which is in Π. Let l be the line connecting the points of G and $\overline{G}$. The perpendicular distance between the point of E and the line l is said to be the *distance* between E and Π.

The set of ∞^2 elements $E(u, v, w)$ which are at a constant distance ρ from a fixed nonlinear flat field $\Pi(\overline{U}, \overline{V}, \overline{W})$ is called a *spherical field* Σ. We term Π the *central flat field* and ρ the *radius* of Σ. The equation of Σ is

$$w = (v - \overline{V}) \cot (u - \overline{U})/2 + \rho \csc (u - \overline{U})/2 - \overline{W}. \tag{13}$$

We call $\Sigma(\overline{U}, \overline{V}, \overline{W}, \rho)$ a set of spherical field coordinates. Obviously

$$\Sigma(\overline{U}, \overline{V}, \overline{W}, \rho) = \Sigma(\overline{U}, \overline{V}, \overline{W}, -\rho).$$

The integral curves of Σ are given in hessian line coordinates by the equation

$$\begin{aligned} v = &- \rho[\cos (u - \overline{U})/2 + \sin^2 (u - \overline{U})/2 \log \cot (u - \overline{U})/4] \\ &- C \cos (u - \overline{U}) + \overline{W} \sin (u - \overline{U}) + \overline{V} + C, \end{aligned} \tag{14}$$

where C is an arbitrary constant. If $\rho = 0$, then Σ becomes its central flat field Π, and its integral curves are the ∞^1 circles which contain the central element $\overline{G}$ of Π. Otherwise if $\rho \neq 0$, the integral curves are transcendental.

Let C be the circular series whose center is the point of the central element $\overline{G}$ of the central flat field Π of the spherical field Σ, whose radius is the radius ρ of Σ, and whose direction is that of $\overline{G}$. We call C the *associated circular series* of Σ. Obviously a spherical field Σ is uniquely determined by its associated circular series C.

The only turbines in a spherical field Σ are the ∞^1 linear turbines whose conjugates are tangent to the associated circular series C of Σ. These are the equiparallel series of Σ. Thus a spherical field Σ contains no nonlinear turbines.

There are ∞^3 limaçon series in a spherical field Σ. Their central turbines are contained in the central flat field Π of Σ. If ρ is the radius of Σ, γ is the radius of any one of these limaçon series L, and α is the angle between Π and the flat field of L, then

$$\rho = 2\gamma \sin \alpha/2. \tag{15}$$

A least limaçon series of a spherical field Σ is any limaçon series of Σ

which has either one of the two equivalent properties: (1) its radius γ is one half the radius ρ of Σ, or (2) its flat field is supplementary or antiparallel to the central flat field Π of Σ. There are ∞^2 least limaçon series in a spherical field Σ.

A spherical field Σ intersects a nonlinear flat field Π_1 which is not parallel to the central flat field Π of Σ in a single limaçon series. If Π_1 and Π are parallel but not identical, then Σ intersects Π_1 in two linear turbines (which may be coincident or imaginary). If Π_1 and Π are identical, then Σ and $\Pi=\Pi_1$ have no common elements. A spherical field intersects a linear flat field in two linear turbines.

Two spherical fields Σ_1 and Σ_2 whose central flat fields Π_1 and Π_2 are not parallel intersect in two limaçon series. If Π_1 and Π_2 are parallel but not identical, then Σ_1 and Σ_2 intersect in four linear turbines (two of which may be coincident, or two or all four of which may be imaginary). If Π_1 and Π_2 are identical, then Σ_1 and Σ_2 have no common elements.

Four elements, at most two of which are parallel and which do not all lie in one flat field, determine *eight* spherical fields. Let us denote the four elements by E_1, E_2, E_3, E_4, where E_3 and E_4 are the two possible parallel ones. Now E_1, E_2, E_j $(j=3, 4)$ determine four limaçon series. The associated circles of these are the inscribed and escribed circles of a triangle T_j which has one vertex at the center O of the turbine determined by E_1 and E_2. Let L_{j1} and L_{j2} denote the two limaçon series whose associated circles are the inscribed circle and the escribed circle opposite the vertex O of T_j. Let L_{j3} and L_{j4} denote the remaining two limaçon series. The central turbines of L_{31}, L_{32}, L_{41}, L_{42} will have a common element F_1, and hence will determine four nonlinear flat fields Π_i $(i=1, 2, 3, 4)$. Similarly, the central turbines of L_{33}, L_{34}, L_{43}, L_{44} will have a common element $F_2 \neq F_1$, and hence will determine four new nonlinear flat fields Π_i $(i=5, 6, 7, 8)$. These eight flat fields Π_i $(i=1, \cdots, 8)$ are the central flat fields of our eight spherical fields Σ_i. The radius ρ_i of Σ_i is the distance between Π_i and any one of the four given elements.

The four elements E_1, E_2, E_3, E_4 such that only E_1, E_2, E_3 are parallel to each other but are not all on one turbine, determine *six* spherical fields. Construct the linear turbine T_1 determined by E_i and E_j $(i, j=1, 2, 3)$. Let Π be the linear flat field which contains the conjugate turbine $\overline{T}_1$ of T_1. Construct the two linear turbines $\overline{T}_2$ and $\overline{T}_3$ contained in Π such that their conjugate turbines T_2 and T_3 contain the remaining two elements. Our spherical fields are those whose associated circular series are tangent to $\overline{T}_1$, $\overline{T}_2$, and $\overline{T}_3$.

The spherical field Σ which has four consecutive elements in common with a general series S at a given element E of S is called the *osculating spherical field* of S at E.

THEOREM 6. *The parameter values of the ∞^1 osculating spherical fields of a general series S are*

$$\begin{gathered}\overline{U} = u - 2 \arctan R'/S', \\ \overline{V} = A\cos\overline{U} + B\sin\overline{U} + R, \qquad \overline{W} = -A\sin\overline{U} + B\cos\overline{U} - S, \\ \rho = \frac{4(r'R' + s'S')}{(R'^2 + S'^2)^{1/2}},\end{gathered} \tag{16}$$

where (a, b, r, s) and (A, B, R, S) are the parameters of the tangent and central turbines of S.

Let v and w in (13) be replaced by functions of u. Upon differentiating this result three times with respect to u and simplifying, we obtain

$$\begin{aligned}(v - \overline{V})\sin(u - \overline{U})/2 + (w + \overline{W})\cos(u - \overline{U})/2 \\ = 2v'\cos(u - \overline{U})/2 - 2w'\sin(u - \overline{U})/2, \\ -\rho/4 = r'\sin(u - \overline{U})/2 + s'\cos(u - \overline{U})/2, \\ 0 = R'\cos(u - \overline{U})/2 - S'\sin(u - \overline{U})/2.\end{aligned} \tag{17}$$

The second and last of these equations give the values of $\overline{U}$ and ρ of (16). Upon replacing ρ in (13) by the value given in the second of these equations, and then solving this result and the first of the above equations for $\overline{V}$ and $\overline{W}$, we find their values to be those of (16). This completes the proof of Theorem 6.

An immediate consequence of Theorem 6 is

THEOREM 7. *The central flat fields of the osculating spherical fields of a general series S are the osculating flat fields of the series of curvature of S (the envelope of the central turbines of S).*

If α denotes the angle between the osculating flat fields of a general series S and its series of curvature, we find from (4) and (16)

$$\sin\alpha/2 = \frac{r'R' + s'S'}{(r'^2 + s'^2)^{1/2}(R'^2 + S'^2)^{1/2}} = \frac{\rho}{2\gamma}, \tag{18}$$

where ρ and γ are the radii of the osculating spherical field Σ and the osculating limaçon series L at an element E of S. Since the central turbine of L is also in the osculating flat field of the series of curvature of S, we find that the following result holds.

THEOREM 8. *The osculating limaçon series L at an element E of a general series S is the intersection of the osculating flat field* Π *and the osculating spherical field* Σ *of S at E.*

From (4), (8), (10), and (11), we find

$$r' + is' = \kappa e^{i(\bar{u}-u)/2}, \qquad R' + iS' = (\kappa\tau - 2i\kappa')e^{i(\bar{u}-u)/2}. \tag{19}$$

Substituting these into the last of equations (16), we obtain

THEOREM 9. *The radius of spherical curvature ρ in terms of the curvature κ and the torsion τ of the general series S is*

$$\rho = \frac{4\kappa^2\tau}{(\kappa^2\tau^2 + 4\kappa'^2)^{1/2}}. \tag{20}$$

Upon substituting the last of equations (19) into the formulas for the curvature and torsion of the series of curvature of a general series S, we obtain the theorem which follows.

THEOREM 10. *The curvature κ_1 and the torsion τ_1 of the series of curvature of a general series S in terms of the curvature κ and the torsion τ of S are*

$$\begin{aligned} \kappa_1 &= (\kappa^2\tau^2 + 4\kappa'^2)^{1/2}, \\ \tau_1 &= \frac{\kappa^2\tau^3 + 4\tau(2\kappa'^2 - \kappa\kappa'') + 4\kappa\kappa'\tau'}{\kappa^2\tau^2 + 4\kappa'^2}. \end{aligned} \tag{21}$$

Differentiating (20) with respect to u, we find

THEOREM 11. *The derivative ρ' of the radius of spherical curvature ρ with respect to u is*

$$\rho' = 4\kappa\kappa'\tau_1/\kappa_1. \tag{22}$$

It may be that the osculating spherical fields of a general series S consist of only one spherical field, namely, the one in which it is contained. In that case, we shall say that S is *cospherical*. From the preceding theorem, we deduce

THEOREM 12. *A general series S is cospherical if and only if its series of curvature is coflat.*

10. **The tangent flat fields of a field.** A set of ∞^2 elements of the plane is called a *field*. We shall omit from consideration the linear flat fields. That is, whenever we speak of a field, we shall understand it to be *not* a linear flat field. It is always possible to find a whirl-motion transformation such that any field F is given by $w = w(u, v)$.

Let $v = v(u)$, $w = w(u, v)$ be a general series S contained in the field F. Its tangent turbines are given by the parameter values

$$\begin{aligned} a &= -w_u \cos u - (\sin u + w_v \cos u)v', \\ b &= -w_u \sin u + (\cos u - w_v \sin u)v', \\ r &= v + w_u + v'w_v, \qquad s = w - v'. \end{aligned} \tag{23}$$

From these equations, we conclude that the following proposition is true.

THEOREM 13. *The tangent turbines of all the series, contained in a field F and passing through an element E of F, constructed at E are contained in a nonlinear flat field.*

The nonlinear flat field of Theorem 13 is called the *tangent flat field* of F at E. Its central element $\overline{E}(\bar{u}, \bar{v}, \bar{w})$ is given by

$$\cos(\bar{u} - u) = -\frac{1 - w_v^2}{1 + w_v^2}, \qquad \sin(\bar{u} - u) = -\frac{2w_v}{1 + w_v^2},$$
$$\bar{v} = v + \frac{2w_u}{1 + w_v^2}, \qquad \bar{w} = -w - \frac{2w_u w_v}{1 + w_v^2}. \tag{24}$$

We call $\overline{E}$ the *conjugate element* of E with respect to F.

Two fields F_1 and F_2 are said to be *tangent* at a common element E if they have the same tangent flat field at E.

As an application of the above, we find that the tangent flat fields of a spherical field Σ consist of the ∞^1 flat fields whose central elements are those of the associated circular series of Σ.

11. **One-parameter families of fields.** The equation

$$w = w(u, v, a) \tag{25}$$

defines a one-parameter family of fields. The series of intersection of any two consecutive fields of this family is called a *characteristic*. The locus of all the characteristics is a field, called the *envelope* of the family. The equations

$$w = w(u, v, a), \qquad w_a(u, v, a) = 0 \tag{26}$$

for each a represents a characteristic of the family. When we eliminate a from the above equations, the result is the equation of the envelope.

It may be easily proved by the preceding equations that *the envelope is tangent to each member of the family at all elements of its characteristic.*

The series of intersection of consecutive characteristics of a one-parameter family of fields is called the *edge of regression*. The eliminants with respect to a of the equations

$$w = w(u, v, a), \qquad w_a(u, v, a) = 0, \qquad w_{aa}(u, v, a) = 0 \tag{27}$$

give the equations of the edge of regression.

We may prove by (26) and (27) that *the edge of regression is tangent to any characteristic at a common element.*

12. **Developable fields.** The envelope of ∞^1 nonlinear flat fields is called a *developable field* F. The series S formed by the central elements of these ∞^1

tangent flat fields of F is called the *associated series* of F. The characteristics of F are turbines. These are called the *generators* of F.

Since each flat field is tangent to the envelope along its characteristic, it follows that *the tangent flat field to a developable field F is the same at all elements of a generator.*

An *umbilical field* F is a developable field whose associated series S is an equiparallel series. Thus a spherical field Σ is an umbilical field whose associated series is a circular series. The generators of an umbilical field F are linear turbines. These are the conjugates of the tangent turbines of its associated equiparallel series S. The edge of regression of F does *not* exist. The equation of any umbilical field F is

$$w = v \cot (u - a)/2 + b(u), \tag{28}$$

where a is a constant.

A developable field F is said to be *general* if its associated series S is a general series. The generators of a general developable field F are nonlinear turbines. These are the tangent turbines of the edge of regression R. Since consecutive generators are the consecutive tangent turbines of R at an element E of R, the osculating flat field of R at E is that flat field of the family which contains these generators. But this flat field is tangent to the developable. Hence *the osculating flat field at any element E of the edge of regression R of a general developable field F is the tangent flat field of F at E.* We find from this that *the edge of regression R and the associated series S of a general developable field F are conjugate series.*

The necessary and sufficient condition that a field $F: w=w(u, v)$ be developable is that its conjugate elements $\overline{E}(\bar{u}, \bar{v}, \bar{w})$ of (24) consist of at most ∞^1 elements. These will then form the associated series S of F. Hence upon setting the three jacobians of the three functions $\bar{u}, \bar{v}, \bar{w}$ of (u, v) equal to zero, we obtain

THEOREM 14. *A field $F: w=w(u, v)$ is a developable field if and only if*

$$(1 + w_v^2 + 2w_{uv})^2 - 4w_{vv}(w_{uu} + w_u w_v) = 0. \tag{29}$$

A field $F: w=w(u, v)$ is a general developable field if and only if the function w of (u, v) satisfies the above equation and $w_{vv} \neq 0$.

13. **Conjugate fields.** The conjugate elements $\overline{E}(\bar{u}, \bar{v}, \bar{w})$ of (24) of a field $F: w=w(u, v)$ form a field $\overline{F}: \bar{w}=\bar{w}(\bar{u}, \bar{v})$ if and only if F is nondevelopable, or if and only if the function w of (u, v) does not satisfy (29). This field $\overline{F}$: $\bar{w}=\bar{w}(\bar{u}, \bar{v})$ is termed the *conjugate field* of F. The equation of $\overline{F}$ is the eliminant with respect to u and v of the equations (24).

Since the conjugate field $\overline{F}$ consists of the central elements of the tangent flat fields of a nondevelopable field F (and conversely), it follows that the tangent turbines of $\overline{F}$ are the conjugates of the tangent turbines of F (and conversely). Hence for the special case when the tangent turbines are circles (the self-conjugate turbines), we deduce the following result.

THEOREM 15. *Two fields F and $\overline{F}$ are conjugate if and only if their integral curves possess the same osculating circles.*

14. **The gaussian curvature of a field.** The series of intersection between a nonlinear flat field and a given field F is called a *flat section* of F. If F is *not* a flat field, there are ∞^3 flat sections in F. There pass ∞^2 flat sections of F through any element E. Finally there are ∞^1 flat sections of F which contain a given element E and which possess a fixed tangent turbine at E.

Let S_1 be any general series contained in a field F. There is a unique flat section S which osculates S_1 at a given element E of S_1. This flat section S is the intersection between the field F and the osculating flat field of S_1 at E. The two series S_1 and S will have the same tangent turbine, the same osculating flat field, the same osculating limaçon series, and the same curvature at E. Thus in order to study the curvatures and the osculating limaçon series of any general series contained in a field F, it is necessary merely to study those of any flat section of F.

Next we shall seek to obtain the curvature κ of any flat section S of a field $F: w=w(u, v)$ at any element E of S in terms of the angle β between the flat field of S and the tangent flat field of F at E. Upon eliminating v'' from the values of r' and s', the first derivatives with respect to u of the last two parameters r and s of the tangent turbine of S at E, we find

$$r' + w_v s' = (w_{uu} + w_u w_v) + (1 + w_v^2 + 2w_{uv})v' + w_{vv}v'^2. \tag{30}$$

We see from (4) and (24) that the angle β satisfies the equation

$$\frac{r'}{s'} = \frac{\sin \beta/2 - w_v \cos \beta/2}{\cos \beta/2 + w_v \sin \beta/2}. \tag{31}$$

Solving the preceding two equations for r' and s', and then substituting these results into the curvature formula, we find that the value of the curvature κ of a flat section S of a field F at any element E of S in terms of the angle β between the flat field of S and the tangent flat field of F at E is

$$\kappa = \frac{(w_{uu} + w_u w_v) + (1 + w_v^2 + 2w_{uv})v' + w_{vv}v'^2}{(1 + w_v^2)^{1/2} \sin \beta/2}. \tag{32}$$

When the angle $\beta=\pi$, we shall call the flat section a *supplementary section* and its curvature at E the *supplementary curvature* κ_s. By the preceding equation, the value of the supplementary curvature κ_s is

$$\kappa_s = \frac{(w_{uu} + w_u w_v) + (1 + w_v^2 + 2w_{uv})v' + w_{vv}v'^2}{(1 + w_v^2)^{1/2}}. \tag{33}$$

By the above two equations, we obtain the following analogue of Meusnier's theorem in the geometry of the whirl-motion group G_6.

THEOREM 16. *Let κ_s and κ be the curvatures of a supplementary section and any other flat section which have the same tangent turbine at a common element E of a field F. If α denotes the angle between these two flat sections, then*

$$\kappa_s = \kappa \cos \alpha/2. \tag{34}$$

The above result shows that a supplementary section possesses the least curvature of all the flat sections of a field F which pass through a given element E in a given tangent turbine direction.

A field F which is generated by a one-parameter family of turbines such that consecutive turbines of the family do not lie in a flat field is termed a *ruled field*. We shall say that a ruled field F is *general* or *special* according as the turbines of the family are nonlinear or linear. A special ruled field F is given by either $w_{vv}=0$, or $w=vm(u)+b(u)$, which is not of the form (28).

A field F: $w=w(u, v)$ is called a *general field* if it is neither a special ruled field nor an umbilical field. Thus F is a general field if and only if $w_{vv}\neq 0$. Of course, the general ruled and the general developable fields are all examples of general fields.

THEOREM 17. *At any element E of a general field F, there is one and only one extremal (maximum or minimum) supplementary curvature κ_0. It is given by the formula*

$$\kappa_0 = \frac{-(1 + w_v^2 + 2w_{uv})^2 + 4w_{vv}(w_{uu} + w_u w_v)}{4w_{vv}(1 + w_v^2)^{1/2}}. \tag{35}$$

For upon completing the square of the quadratic expression in v' of (33), we find that (33) may be written in the form

$$\kappa_s = \frac{-(1 + w_v^2 + 2w_{uv})^2 + 4w_{vv}(w_{uu} + w_u w_v)}{4w_{vv}(1 + w_v^2)^{1/2}} + \frac{[(1 + w_v^2 + 2w_{uv}) + 2w_{vv}v']^2}{4w_{vv}(1 + w_v^2)^{1/2}}. \tag{36}$$

This will be a maximum or a minimum with respect to v' only when the squared bracket is zero. The remaining part of the above expression will give us the value of κ_0 of (35). Theorem 17 is completely proved.

Through any element E of a field F, there passes a unique equiparallel series S contained in F. The curvature of S at E is called the *equiparallel curvature* λ of F at E. It is given by

$$\lambda = \frac{w_{vv}}{(1 + w_v^2)^{3/2}}. \tag{37}$$

By means of (35) and (37), we find that (33) or (36) may be written in the form

$$\kappa_s = \kappa_0 + \lambda(1 + w_v^2)\left[v' + \frac{1 + w_v^2 + 2w_{uv}}{2w_{vv}}\right]^2. \tag{38}$$

If δ denotes the distance between the centers of the tangent turbines of (23) of the extremal supplementary section and any supplementary section, we obtain the following analogue of Euler's theorem.

THEOREM 18. *Let κ_0 be the extremal supplementary curvature and λ the equiparallel curvature at an element E of a general field F. If δ is the distance between the centers of the tangent turbines of the extremal supplementary section and any supplementary section whose curvature at E is κ_s, then*

$$\kappa_s = \kappa_0 + \lambda\delta^2. \tag{39}$$

The *gaussian curvature* K of any field F at any element E of F is given by the formula

$$K = \frac{(1 + w_v^2 + 2w_{uv})^2 - 4w_{vv}(w_{uu} + w_u w_v)}{(1 + w_v^2)^2}. \tag{40}$$

We note that K is zero if and only if F is developable.

THEOREM 19. *The gaussian curvature K at any element E of a general field F is minus four times the product of the extremal supplementary curvature κ_0 and the equiparallel curvature λ at E. That is*

$$K = -4\kappa_0\lambda. \tag{41}$$

By relations (23), (33), and (40), we now deduce the following proposition.

THEOREM 20. *If δ is the distance between the centers of the tangent turbines of any two supplementary sections whose curvatures are κ_s and κ_s' at a common element E of a special ruled field F, then*

$$\kappa_s = \kappa_s' + K^{1/2}\delta. \tag{42}$$

For an umbilical field F of (28), all the supplementary curvatures at an element E of F are equal, and they have the common value

$$\kappa_s = \frac{w_{uu} + w_u w_v}{(1 + w_v^2)^{1/2}} = b'' \sin (u - a)/2 + b' \cos (u - a)/2. \tag{43}$$

Through any element E of a general field F, there is a single tangent turbine direction which gives the extremal supplementary curvature κ_0. This is called the *principal direction* of F at E. Any general series S of a general field F such that the tangent turbine direction of any element E of S is a principal direction is called a *principal series*. The differential equation of all principal series of a general field F is

$$2w_{vv}v' + (1 + w_v^2 + 2w_{uv}) = 0. \tag{44}$$

Since the equiparallel curvature λ at an element E of any field F is a sort of extremal curvature, the equiparallel series of F may be considered to be principal series. Thus a general field F possesses $2\infty^1$ principal series, namely, (1) the ∞^1 general series which satisfy (44), and (2) the ∞^1 equiparallel series. A special ruled field F has only ∞^1 principal series, the equiparallel series of F. Finally all ∞^∞ series of an umbilical field F are principal series.

The angle α between the two tangent flat fields at a common element E of two fields $F: w = f(u, v)$ and $G: w = g(u, v)$ is defined to be the angle between F and G at E. By (24), the derivative of this angle α with respect to u is

$$\alpha' = \frac{(1 + f_v^2 + 2f_{uv}) + 2f_{vv}v'}{1 + f_v^2} - \frac{(1 + g_v^2 + 2g_{uv}) + 2g_{vv}v'}{1 + g_v^2}. \tag{45}$$

We deduce from this the following analogue of Joachimsthal's theorem.

THEOREM 21. *If the series of intersection of two fields is a principal series on both, the fields cut at a constant angle. Conversely, if two fields cut at a constant angle, and the series of intersection is a principal series on one, then it is a principal series on the other.*

15. **The gaussian curvature of the conjugate field.** From (24), we find that the partial derivatives of the first and second orders with respect to $\bar{u}$ and $\bar{v}$ of the function $\bar{w} = \bar{w}(\bar{u}, \bar{v})$ which defines the field $\overline{F}$ conjugate to the field $F: w = w(u, v)$ are

$$
\begin{aligned}
&\frac{\partial \bar{w}}{\partial \bar{u}} = -w_u, \qquad \frac{\partial \bar{w}}{\partial \bar{v}} = -w_v,\\
&\frac{\partial^2 \bar{w}}{\partial \bar{u}^2} = \frac{-w_{uu}(1+w_v^2)^2 + 4w_u w_v(w_{uu}w_{vv} - w_{uv}^2)}{(1+w_v^2+2w_{uv})^2 - 4w_{vv}(w_{uu}+w_u w_v)},\\
&\frac{\partial^2 \bar{w}}{\partial \bar{u}\partial \bar{v}} = \frac{-w_{uv}(1+w_v^2)^2 + 2(1+w_v^2)(w_{uu}w_{vv} - w_{uv}^2)}{(1+w_v^2+2w_{uv})^2 - 4w_{vv}(w_{uu}+w_u w_v)},\\
&\frac{\partial^2 \bar{w}}{\partial \bar{v}^2} = \frac{-w_{vv}(1+w_v^2)^2}{(1+w_v^2+2w_{uv})^2 - 4w_{vv}(w_{uu}+w_u w_v)}.
\end{aligned} \tag{46}
$$

By these equations, we deduce the following result.

THEOREM 22. *The product of the gaussian curvatures K and $\overline{K}$ at conjugate elements E and $\overline{E}$ of two conjugate fields F and $\overline{F}$ is unity. That is*

$$K\overline{K} = 1. \tag{47}$$

The series S of a field F and the series $\overline{S}$ of the field $\overline{F}$ conjugate to F are said to be conjugate with respect to F or $\overline{F}$ if they are corresponding series under the transformation (24).

For any two series S and $\overline{S}$ conjugate with respect to two conjugate general fields F or $\overline{F}$, we obtain the following relation:

$$1 + \bar{w}_{\bar{v}}^2 + 2\bar{w}_{\bar{u}\bar{v}} + 2\bar{w}_{\bar{v}\bar{v}}\frac{d\bar{v}}{d\bar{u}} = \frac{(1+w_v^2)^2}{1+w_v^2+2w_{uv}+2w_{vv}v'}. \tag{48}$$

By substituting this into the supplementary curvature formula of $\overline{F}$, we prove

THEOREM 23. *Let E be any element of a general nondevelopable field F and $\overline{E}$ its conjugate element of the general field $\overline{F}$ conjugate to F. The corresponding supplementary curvatures κ_s and $\bar{\kappa}_s$, the extremal supplementary curvatures κ_0 and $\bar{\kappa}_0$, and the equiparallel curvatures λ and $\bar{\lambda}$ of F and $\overline{F}$ at E and $\overline{E}$ are related by the formulas*

$$\bar{\kappa}_s = \frac{\kappa_s}{4\lambda(\kappa_s - \kappa_0)}, \qquad \bar{\kappa}_0 = \frac{1}{4\lambda}, \qquad \bar{\lambda} = \frac{1}{4\kappa_0}. \tag{49}$$

For any two series S and $\overline{S}$ conjugate with respect to two conjugate special ruled fields F or $\overline{F}$, we find the following relation (since $w_{vv}=0$)

$$\frac{d\bar{v}}{d\bar{u}} = v' + \frac{2(1+w_v^2)w_{uu} - 4w_u w_v w_{uv}}{(1+w_v^2)(1+w_v^2+2w_{uv})}. \tag{50}$$

Upon substituting this into the supplementary curvature formula of $\overline{F}$, we obtain the following result.

THEOREM 24. *Let E be any element of a special ruled field F and $\overline{E}$ its conjugate element of the special ruled field $\overline{F}$ conjugate to F. The corresponding supplementary curvatures κ_s and $\bar{\kappa}_s$ of F and $\overline{F}$ at E and $\overline{E}$ are related by the formula*

$$\bar{\kappa}_s = \frac{\kappa_s}{K}. \tag{51}$$

The principal series of two conjugate fields F and $\overline{F}$ are conjugate with respect to F or $\overline{F}$. A principal general series of one of these two general fields corresponds by the transformation (24) to an equiparallel series of the other. Otherwise the equiparallel series of these two special ruled fields correspond to each other under the transformation (24).

16. **The osculating limaçon series of a field.** Upon solving the equations (30) and (31) for r' and s' and making use of the supplementary curvature formula (33), we find that their values are

$$r' = \frac{\kappa_s(\sin\beta/2 - w_v\cos\beta/2)}{(1+w_v^2)^{1/2}\sin\beta/2}, \qquad s' = \frac{\kappa_s(\cos\beta/2 + w_v\sin\beta/2)}{(1+w_v^2)^{1/2}\sin\beta/2}. \tag{52}$$

Substituting these values into (8), we see that the parameters of the osculating limaçon series L of any flat section S of a field $F: w=w(u,v)$ at an element E of F are

$$\begin{aligned}
A &= a + \frac{2\kappa_s}{(1+w_v^2)^{1/2}\sin\beta/2}[\cos(\beta/2-u) + w_v\sin(\beta/2-u)],\\
B &= b + \frac{2\kappa_s}{(1+w_v^2)^{1/2}\sin\beta/2}[-\sin(\beta/2-u) + w_v\cos(\beta/2-u)],\\
C &= -\frac{4\kappa_s}{(1+w_v^2)^{1/2}\sin\beta/2}[\cos(\beta-u)/2 + w_v\sin(\beta-u)/2],\\
D &= \frac{4\kappa_s}{(1+w_v^2)^{1/2}\sin\beta/2}[\sin(\beta-u)/2 - w_v\cos(\beta-u)/2],\\
R &= r + \frac{2\kappa_s}{(1+w_v^2)^{1/2}\sin\beta/2}[\cos\beta/2 + w_v\sin\beta/2],\\
S &= s - \frac{2\kappa_s}{(1+w_v^2)^{1/2}\sin\beta/2}[\sin\beta/2 - w_v\cos\beta/2],
\end{aligned} \tag{53}$$

where (a, b, r, s) are the parameters of the tangent turbine T of (23) of S at E,

κ_s is the supplementary curvature at E of the supplementary section of F which is tangent to S at E, and β is the angle between the flat field of S and the tangent flat field of F at E.

Next let us consider all the ∞^1 flat sections of the field F which pass through the element E of F and which possess the same tangent turbine T at E. In the formulas (53) for the ∞^1 osculating limaçon series of these flat sections at E, we observe that *only* the angle β is variable. The ∞^1 central turbines of these osculating limaçon series all contain the element E_1 of the tangent turbine T which is supplementary (antiparallel) to E. By (53), it may be proved after some calculation that these central turbines all are contained in the flat field II whose central element $\overline{G}(U, V, W)$ is given by

$$\cos(U-u) = -\frac{1-w_v^2}{1+w_v^2}, \qquad \sin(U-u) = -\frac{2w_v}{1+w_v^2},$$

(54)

$$V = v + \frac{2w_u}{1+w_v^2} + \frac{4\kappa_s w_v}{(1+w_v^2)^{1/2}}, \qquad W = -w - \frac{2w_u w_v}{1+w_v^2} + \frac{4\kappa_s}{(1+w_v^2)^{1/2}}.$$

From these equations and from Meusnier's Theorem 16 (formula 34), we deduce

THEOREM 25. *Let us consider the ∞^1 flat sections of a field F which pass through an element E of F and which possess the same tangent turbine T at E. The ∞^1 osculating limaçon series of these flat sections at E generate a spherical field Σ.*

Let E_1 be the element on the tangent turbine T which is supplementary (antiparallel) to the fixed element E. Let T_S be the linear turbine whose direction is that of the central element $\overline{E}$ of the tangent flat field of F at E, and whose base line joins the points of $\overline{E}$ and E_1. The central element $\overline{G}$ of the central flat field II of the spherical field Σ of Theorem 25 is on the linear turbine T_S and the distance of $\overline{G}$ from $\overline{E}$ is $4\kappa_s$. The radius ρ of Σ is also $4\kappa_s$.

If we vary the tangent turbine direction T of Theorem 25, there will result ∞^1 spherical fields. The ∞^1 central elements of their central flat fields will generate the linear turbine T_S.

17. **The geodesic series of a field.** A series S of a field F is termed a *geodesic series* if its curvature at any element E of S does not exceed the curvature at E of any other series of F which is tangent to S at E. By setting the partial derivative with respect to v'' of the curvature κ of any general series of a field F equal to zero, and solving the result for v'', we find that the differential equation of all the geodesic series of a field F: $w = w(u, v)$ is

$$v'' = \frac{w_u - w_v w_{uu} - 2w_v w_{uv} v' - w_v w_{vv} v'^2}{1 + w_v^2}. \tag{55}$$

There are ∞^2 geodesic series, all general, in a given field F. The ∞^1 equiparallel series of a field F are also considered to be geodesic series. The geodesic series of a spherical field are its ∞^2 least limaçon series, together with its ∞^1 linear turbines. The geodesic series of a flat field are its ∞^2 turbines.

The curvature κ of any geodesic series S of a field F at any element E of S is equal to the supplementary curvature κ_s of the supplementary section of F which is tangent to S at E.

To calculate the torsion τ of a geodesic series S, we proceed as follows. The derivatives r' and s' with respect to u of the last two parameters r and s of the tangent turbine T of S at any element E of S are given by (52) where $\beta = \pi$. Thus $s'/r' = w_v$. By this and (4), we find that the normal angle $\bar{u}$ of the osculating flat field of S at E is

$$\bar{u} = u + 2 \arctan w_v. \tag{56}$$

Differentiating this with respect to u, we see that the torsion τ of any geodesic series S at any element E of S is

$$\tau = \frac{1 + w_v^2 + 2w_{uv} + 2w_{vv} v'}{1 + w_v^2}. \tag{57}$$

THEOREM 26. *The torsion τ of a geodesic series S of a field F at an element E of S is zero if and only if S is tangent to a principal series of F at E. The necessary and sufficient condition that a geodesic series be a principal series is that it be coflat.*

By equation (57), we obtain the following two results.

THEOREM 27. *Let λ be the equiparallel curvature of a general field F at an element E of F. Let δ be the distance between the centers of the tangent turbines of the extremal supplementary section and any geodesic series S through E. The torsion τ of S is*

$$\tau = 2\lambda\delta. \tag{58}$$

THEOREM 28. *The torsion τ of any geodesic series S of a special ruled or an umbilical field F at any element E of S is*

$$\tau = K^{1/2}. \tag{59}$$

To define the *geodesic curvature* κ_g of any series S of a field F at any element E of S, we proceed as follows. Let S_g be the geodesic series tangent to S

at E. Let E_s be the element on S which makes an angle Δu with E, and let E_g be the element on S_g which makes the same angle Δu with E. At E_s and E_g, construct the two tangent turbines T_s and T_g of S and S_g. Let $\Delta\delta$ be the distance between the centers of T_s and T_g. Then

$$\kappa_g = \frac{d\delta}{du} = \lim_{\Delta u \to 0} \frac{\Delta\delta}{\Delta u} \tag{60}$$

is defined to be the geodesic curvature of S at E.

It is found that $\kappa_g = (1+w_v^2)^{1/2}(v''-v_g'')$, where $\iota'_g{}'$ belongs to the geodesic series S_g. From this, it follows that the geodesic curvature κ_g of any series S of a field F: $w=w(u, v)$ at any element E of S is

$$\kappa_g = \frac{(1+w_v^2)v'' - (w_u - w_v w_{uu}) + 2w_v w_{uv} v' + w_v w_{vv} v'^2}{(1+w_v^2)^{1/2}}. \tag{61}$$

By means of (33), (52), and (61), we obtain the result which follows.

THEOREM 29. *Let β be the angle between the osculating flat field of a general series S of a field F and the tangent flat field of F at an element E of S. Let κ be the curvature and κ_g the geodesic curvature of S at E. Let κ_s be the supplementary curvature of the supplementary section which is tangent to S at E. Then*

$$\kappa_s = \kappa \sin \beta/2 = -\kappa_g \tan \beta/2, \quad \kappa_g = -\kappa \cos \beta/2, \quad \kappa_s^2 + \kappa_g^2 = \kappa^2. \tag{62}$$

18. **The asymptotic series of a field.** Reciprocal directions at an element E of a field F may be defined as follows. Let G be an element in F adjacent to E. Let ER be the turbine of intersection of the tangent flat fields of F at E and G. As G tends to coincidence with E, the limiting tangent turbine directions of EG and ER are said to be *reciprocal* at E.

The necessary and sufficient condition that the tangent turbine directions dv/du and $\delta u/\delta v$ be reciprocal are

$$2w_{vv}\frac{\delta v}{\delta u}\frac{dv}{du} + (1 + w_v^2 + 2w_{uv})\left(\frac{\delta v}{\delta u} + \frac{dv}{du}\right) + 2(w_{uu} + w_u w_v) = 0. \tag{63}$$

THEOREM 30. *Let T_0 be the principal tangent turbine (the principal direction), and let T and T_1 be any other two tangent turbines at an element E of a general field F. Let δ and δ_1 be the distances of the centers of T and T_1 from that of T_0. The directions of T and T_1 are reciprocal if and only if*

$$\delta\delta_1 = -\kappa_0/\lambda, \tag{64}$$

where κ_0 is the extremal supplementary curvature and λ is the equiparallel curvature of F at E.

THEOREM 31. *The two tangent turbine directions T and T_1 at an element E of a special ruled field F are reciprocal if and only if the sum of the supplementary curvatures κ_s and κ_s' of T and T_1 is zero. That is*

$$\kappa_s + \kappa_s' = 0. \tag{65}$$

Any two tangent turbine directions at an element E of an umbilical field F are reciprocal.

In general, given a one-parameter family of series $\phi(u, v) = \text{const.}$ of a general or special ruled field F, we can find another one-parameter family of series $\psi(u, v) = \text{const.}$ of F such that the two tangent turbine directions of the two series of the two families passing through any element E of F are reciprocal at E.

The self-reciprocal directions of a field F are called *asymptotic directions*. Any series S of a field F whose tangent turbine direction at any element E of S is an asymptotic direction is termed an *asymptotic series*. The differential equation of all asymptotic series of a field F is

$$w_{vv}v'^2 + (1 + w_v^2 + 2w_{uv})v' + (w_{uu} + w_u w_v) = 0. \tag{66}$$

This means that a series S is an asymptotic series if and only if the supplementary curvature κ_s of the supplementary section tangent to S at any element E of S is zero at E.

In a general field F, there are $2\infty^1$ asymptotic series, all general series. A special ruled field F possesses $2\infty^1$ asymptotic series, namely, (1) the ∞^1 general series which satisfy (66), and (2) the ∞^1 equiparallel series of F. Every series of an umbilical field F is an asymptotic series.

From Theorem 30 we pass to the following conclusion.

THEOREM 32. *Let T_0 be the principal tangent turbine (the principal direction), and let T be any other tangent turbine at an element E of a general field F. Let δ be the distance between the centers of T_0 and T. The tangent turbine direction T is an asymptotic direction if and only if*

$$\delta = (-\kappa_0/\lambda)^{1/2}. \tag{67}$$

The osculating flat field of a general series S of a general or special ruled field F at any element E of S will coincide with the tangent flat field of F at E if and only if the normal angles of the central elements of these two flat fields are identical. This means that the angle of equations (52) must be zero. Hence the supplementary curvature κ_s of the supplementary section tangent to S at E is zero at E. Therefore S is an asymptotic series. Thus we obtain

THEOREM 33. *A general series S of a general or special ruled field F is an*

asymptotic series of F if and only if its osculating flat fields coincide with the tangent flat fields of F at the elements of S.

By this result and by (4) and (24), it follows that the torsion τ of an asymptotic series S at any element E of S is the same as that of the geodesic series which is tangent to S at E. Hence τ is given by (57). Upon squaring this value of τ and noting that S satisfies (66), we obtain the following analogue of the Beltrami-Enneper theorem.

THEOREM 34. *The torsion τ of any asymptotic general series S of a general or special ruled field F at any element E of S is equal to the square root of the gaussian curvature K of F at E. That is,*

$$\tau = K^{1/2}. \tag{68}$$

Thus we have discussed in the geometry of the whirl-motion group G_6 the analogues of some of the classic theorems in the differential geometry of curves and surfaces embedded in a euclidean three-dimensional space. Of significant interest is the fact that our geometric configurations and invariants may be constructed by ordinary geometric means. Some of our results which seem to be completely analogous in content are nevertheless entirely distinct when we think of the meanings of the terms used.

BIBLIOGRAPHY

1. Kasner, *The group of turns and slides, and the geometry of turbines*, American Journal of Mathematics, vol. 33 (1911), pp. 193–202.

2. Kasner and De Cicco, *The geometry of turbines, flat fields, and differential equations*, American Journal of Mathematics, vol. 59 (1937), pp. 545–563.

3. De Cicco, *The geometry of whirl series*, these Transactions, vol. 43 (1938), pp. 344–358.

4. Kasner and De Cicco, *The geometry of the whirl-motion group G_6: elementary invariants*, Bulletin of the American Mathematical Society, vol. 44 (1938), pp. 399–403.

5. Kasner and De Cicco, *Quadric fields in the geometry of the whirl-motion group G_6*, American Journal of Mathematics, vol. 61 (1939), pp. 131–142.

6. De Cicco, *The differential geometry of series of lineal elements*, these Transactions, vol. 46 (1939), pp. 348–361.

BROOKLYN COLLEGE,
BROOKLYN, N. Y.

THE ROLE OF THE MEAN CURVATURE IN THE IMMERSION THEORY OF SURFACES*

BY

H. W. ALEXANDER

1. **Introduction.** In this paper it is proposed to deal with the immersion theory of surfaces from a point of view somewhat different from the classical. Part I, sections 3–6, will be devoted to an exposition of the role of the mean curvature in the immersion theory. In the general case, expressions will be obtained for the second fundamental tensor $\Omega_{\alpha\beta}$ in terms of the mean curvature, the first fundamental tensor and their derivatives; and necessary and sufficient conditions will be derived in order that a function $K_m(u^\alpha)$ may constitute the mean curvature of a surface with given linear element. Given a function $K_m(u^\alpha)$ satisfying these conditions, the surface will be determined to within rigid motions in space, except in certain particular cases, which will be given careful treatment.

The dependence of the immersion on the mean curvature is considered in two papers by W. C. Graustein [1], in the first of which ample references to the literature are given. The expression for $\Omega_{\alpha\beta}$ in terms of the mean curvature, and the differential equations governing the mean curvature, do not appear to have been obtained before.

Part II, sections 7–11, will deal with an important type of singularity of the immersion, referred to as the edge of regression, which is the envelope of both sets of asymptotic lines, and for which the mean curvature is infinite. The geometrical properties of the edge, both local and in the large, as well as its analytical structure, will be studied in these sections.

PART I

2. **Notation and definition.** In Part I we shall be dealing purely with the local geometry of a surface; that is, with a region R of the surface which is homeomorphic to the interior of a circle of the Euclidean plane. The coordinates u^α, $\alpha = 1, 2$, of a point of R are defined to be the Cartesian coordinates of its image under the homeomorphism, and the homeomorphism defines a coordinate system over R, denoted by $[u^\alpha]$. The functions we shall consider are assumed to possess a sufficient number of derivatives so that the quantities defined shall exist.

* Presented to the Society, in part on September 9, 1937, and in part on December 28, 1939; received by the editors July 12, 1939, and, in revised form, November 18, 1939.

The region R belongs to an immersed surface if to each point of R there corresponds an ordered triad of numbers x^i, $i=1, 2, 3$,* called its space coordinates, such that the sum of the squares of the Jacobians of the x's with respect to the u's is not identically zero.

The region R belongs to an intrinsic surface if to each coordinate system there corresponds an ordered triad of functions $g_{11}(u^\alpha)$, $g_{12}(u^\alpha)=g_{21}(u^\alpha)$, $g_{22}(u^\alpha)$, called the components of its first fundamental tensor, such that $g=|g_{\alpha\beta}|\neq 0$, $g_{\alpha\beta}du^\alpha du^\beta$ is positive definite, and $g_{\alpha\beta}$ undergoes the transformation

$$\bar{g}_{\alpha\beta}=g_{\gamma\delta}\frac{\partial u^\gamma}{\partial \bar{u}^\alpha}\frac{\partial u^\delta}{\partial \bar{u}^\beta}$$

when the coordinate system $[u^\alpha]$ is transformed into the coordinate system $[\bar{u}^\alpha]$. The surface is regular at a point P if there exists a coordinate system for which $g\neq 0$ at P. We shall deal only with such surfaces.

From an immersed surface we can obtain an intrinsic surface by defining the fundamental tensor

$$g_{\alpha\beta}=\frac{\partial x^i}{\partial u^\alpha}\cdot\frac{\partial x^i}{\partial u^\beta}; \tag{2.1}$$

conversely, given an intrinsic surface by means of the functions $g_{\alpha\beta}(u^\gamma)$, we may be able to integrate (2.1) to obtain an immersed surface.

If ε_{ijk} denotes a factor which is zero unless i, j, k are different, and 1 or -1 according as ijk are in cyclic or anticyclic order, and if $\mathfrak{e}^{\alpha\beta}$ is the skew symmetric tensor whose components are [2]

$$\mathfrak{e}^{11}=\mathfrak{e}^{22}=0, \qquad \mathfrak{e}^{12}=-\mathfrak{e}^{21}=g^{-1/2},$$

then the second fundamental tensor is given by

$$\Omega_{\alpha\beta}=\varepsilon_{ijk}\mathfrak{e}^{\gamma\delta}\frac{\partial x^i}{\partial u^\gamma}\frac{\partial x^j}{\partial u^\delta}\frac{\partial^2 x^k}{\partial u^\alpha \partial u^\beta},$$

and the mean curvature by

$$K_m=g^{\alpha\beta}\Omega_{\alpha\beta}.$$

The quantities $\Omega_{\alpha\beta}$ satisfy the equations of Codazzi

$$\Omega_{\alpha\beta,\gamma}-\Omega_{\alpha\gamma,\beta}=0, \tag{2.2}$$

where the comma indicates covariant differentiation, and satisfy the equations of Gauss

* We shall employ Greek indices to denote the range 1, 2, and Latin indices to denote the range 1, 2, 3.

$$Kg = \Omega,$$

where Ω is the determinant of $\Omega_{\alpha\beta}$, and K is the Gaussian curvature, which is expressible in terms of the $g_{\alpha\beta}$ and their derivatives.

In the classical treatment, the immersion theory of surfaces is broken down into two stages: first, the discovery of a tensor $\Omega_{\alpha\beta}$ satisfying the equations of Gauss and Codazzi; and, second, the integration of certain equations of the Riccati type to obtain the space coordinates of the surface. The theorem is proved that if an intrinsic surface is initially given, and if there exist functions $\Omega_{\alpha\beta}(u^\gamma)$ satisfying the equations of Gauss and Codazzi, then there exists an immersion of the surface, unique to within rigid motions in space, for which $\Omega_{\alpha\beta}(u^\gamma)$ is the second fundamental tensor. In the present paper it is proposed to introduce a stage prior to the two mentioned: namely, to discover a single function K_m (the mean curvature) satisfying, in general, two partial differential equations of the third order. The tensor $\Omega_{\alpha\beta}$ may then be obtained without integration.

We shall assume that the region R is free from umbilics. That is, if ρ_1 and ρ_2 are the principal radii of curvature, $\rho_1 \neq \rho_2$ over R. Let us choose ρ_2 to be the larger of the two radii, and define

$$M = \frac{K_m}{2} = (1/2)\left(\frac{1}{\rho_1} + \frac{1}{\rho_2}\right), \qquad N = (M^2 - K)^{1/2} = (1/2)\left(\frac{1}{\rho_1} - \frac{1}{\rho_2}\right).$$

As a notational convenience we introduce the idea of the "bisector tensor" of a real symmetric tensor. The symmetric tensor $a^*_{\alpha\beta}$ is the bisector tensor of the symmetric tensor $a_{\alpha\beta}$ if $a^*_{\alpha\beta}du^\alpha du^\beta$ is the Jacobian of $a_{\alpha\beta}du^\alpha du^\beta$ and $g_{\alpha\beta}du^\alpha du^\beta$ divided by $4g^{1/2}$. It may be shown that

$$a^*_{\alpha\beta} = (1/2)(\varepsilon^{\delta}_{\cdot\alpha}a_{\delta\beta} + \varepsilon^{\delta}_{\cdot\beta}a_{\delta\alpha}),$$

where $\varepsilon^{\delta}_{\cdot\alpha} = g_{\gamma\alpha}\varepsilon^{\delta\gamma}$. By definition the directions $a^*_{\alpha\beta}du^\alpha du^\beta = 0$ bisect the angles between the directions $a_{\alpha\beta}du^\alpha du^\beta = 0$. Thus the equations $\Omega_{\alpha\beta}du^\alpha du^\beta = 0$ and $\Omega^*_{\alpha\beta}du^\alpha du^\beta = 0$ define, respectively, the asymptotic and principal directions.

The bisector tensor satisfies the identity

$$a^* = |a^*_{\alpha\beta}| = a - (g/4)(g^{\alpha\beta}a_{\alpha\beta})^2.$$

The following three conditions, A, B and C, are equivalent to each other:

$$(2.3) \qquad \text{A: } a^* = 0; \qquad \text{B: } \frac{a_{11}}{g_{11}} = \frac{a_{12}}{g_{12}} = \frac{a_{22}}{g_{22}}; \qquad \text{C: } a^*_{\alpha\beta} = 0.$$

3. **General expressions for $\Omega_{\alpha\beta}$.** In §§3 and 4 we shall assume we are given an immersed surface, and we shall obtain necessary conditions on the mean curvature, and on an invariant ϕ to be defined.

THEOREM 3.1. *Let $A(u^\alpha)$ be any real non-constant scalar (that is, $\Delta_1 A \neq 0$),† and let $b_{\alpha\beta}$ be the tensor $(2NA_{,\alpha}A_{,\beta})/\Delta_1 A$. Let ϕ be twice the angle from the vector $A_{,\alpha}$ to the line of curvature whose principal radius is ρ_1. Then*

$$\Omega_{\alpha\beta} = \cos\phi(b_{\alpha\beta} - Ng_{\alpha\beta}) + \sin\phi b^*_{\alpha\beta} + Mg_{\alpha\beta}, \tag{3.1}$$

where $b^*_{\alpha\beta}$ is the bisector tensor of $b_{\alpha\beta}$.

Proof. Use the coordinate system consisting of the lines $A = \text{const.}$ and their orthogonal trajectories, with $u^1 = A$. In this system equations (3.1) reduce to

$$\Omega_{11} = g_{11}(N\cos\phi + M),\quad \Omega_{12} = g^{1/2}N\sin\phi,\quad \Omega_{22} = g_{22}(-N\cos\phi + M). \tag{3.2}$$

If θ is the angle between the direction du^α and the line curvature whose principal radius is ρ_1, then from Euler's theorem the normal curvature corresponding to the direction du^α is given by

$$\frac{\Omega_{\alpha\beta}du^\alpha du^\beta}{g_{\alpha\beta}du^\alpha du^\beta} = N\cos 2\theta + M. \tag{3.3}$$

Using (3.3) with $du^2 = 0$, $2\theta = \phi$, we obtain the expression for Ω_{11} given in (3.2). Using (3.3) with $du^1 = 0$ and $2\theta = \phi + \pi$, we obtain Ω_{22}. Finally, to obtain Ω_{12} we use

$$\Omega_{12}^2 = \Omega_{11}\Omega_{22} - Kg = gN^2\sin^2\phi.$$

Thus equation (3.2) is verified.

THEOREM 3.2. *The invariant ϕ satisfies the differential equation*

$$\phi_{,\alpha} = \frac{1}{N\Delta_1 A}\left[2A_{,\gamma}\Delta_1(M, A) - M_{,\gamma}\Delta_1 A\right]\cdot\left[\varepsilon^{\gamma}_{\cdot\alpha}\cos\phi - \delta^\gamma_\alpha\sin\phi\right] - \frac{2\varepsilon^{\beta\gamma}A_{,\beta}A_{,\gamma\alpha}}{\Delta_1 A} - \frac{\varepsilon^{\gamma}_{\cdot\alpha}N_{,\gamma}}{N}, \tag{3.4}$$

where δ^γ_α is the Kronecker delta.

By using the coordinate system of the proof of Theorem 3.1, it is readily verified that equations (3.4) are equivalent to the Codazzi equations (2.2), where $\Omega_{\alpha\beta}$ is given by (3.1).

† Throughout the paper the signs $=$ and $\neq$ will be assumed to hold at every point of the region R.

4. **Necessary conditions on the mean curvature.** From this point on we shall specialize the invariant A to equal either M or K. If M is not constant, we put $A=M$; if M is constant, but K is not, we put $A=K$. It is easily proved that if both M and K are constant, the surface reduces to a plane, a sphere or a circular cylinder.

We consider first the case $\Delta_1 M \neq 0$. Then we may take $A=M$, and (3.4) reduces to

$$\phi_{,\alpha} = \frac{M_{,\gamma}}{N}\left(\varepsilon^{\gamma}_{\cdot\alpha}\cos\phi - \delta^{\gamma}_{\alpha}\sin\phi\right) - \frac{2\varepsilon^{\beta\gamma}M_{,\beta}M_{,\gamma\alpha}}{\Delta_1 M} - \varepsilon^{\gamma}_{\cdot\alpha}\frac{N_{,\gamma}}{N}. \tag{4.1}$$

THEOREM 4.1. *The invariants M and ϕ satisfy the equation*

$$R\cos\phi + S\sin\phi + T = 0, \tag{4.2}$$

where†

$$\begin{aligned} R &= (1/N)\left[\Delta_2 M - \Delta_1(M, \log \Delta_1 M/N^2)\right], \\ S &= (-1/N)\Theta(M, \log \Delta_1 M/N^2), \\ T &= \Delta_2 \log N - \Delta_1 M/N^2 - 2K. \end{aligned} \tag{4.3}$$

Equation (4.2) is the condition of integrability of equations (4.1). That this is the case may be verified by using the coordinate system of the proof of Theorem 3.1 with $A=M$.

The invariant T of (4.3) may also be expressed in the form

$$T = \frac{-M}{N^2}g^{\alpha\beta}a_{\alpha\beta} - \frac{\Delta_2 K}{2N^2} - \frac{\Delta_1 K}{2N^4} - 2K, \tag{4.4}$$

where

$$a_{\alpha\beta} = -M_{,\alpha\beta} + (1/N)(M_{,\alpha}N_{,\beta} + M_{,\beta}N_{,\alpha}). \tag{4.5}$$

This expression is obtained from those of (4.3) by expanding $\Delta_2 \log N = (1/2)\Delta_2 \log (M^2-K)$.

If $\Omega^{*\alpha\beta}$ is the cofactor of $\Omega^{*}_{\alpha\beta}$ in the determinant $|\Omega^{*}_{\alpha\beta}|$ divided by $\Omega^{*}=-gN^2$, and if P and Q are given by $P=-\Omega^{*\alpha\beta}a_{\alpha\beta}$, $Q=-\Omega^{*\alpha\beta}a^{*}_{\alpha\beta}$, where $a_{\alpha\beta}$ is given by (4.5), then P and Q satisfy the identities

$$\begin{aligned} P &= -S\cos\phi + R\sin\phi, \qquad Q = -R\cos\phi - S\sin\phi, \\ P^2 + Q^2 &= R^2 + S^2 = -4a^{*}/gN^2. \end{aligned} \tag{4.6}$$

The proof follows readily when the coordinate system of Theorem 3.1, with $A=M$, is employed.

† $\Theta(A, B)$ is defined by $\Theta(A, B)=\varepsilon^{\alpha\beta}A_{,\alpha}B_{,\beta}$.

From (4.2) and (4.6) we have $Q=T$; and hence, using (4.6) it follows that P and Q are both expressible in terms of M and $g_{\alpha\beta}$ and their derivatives:

$$P = \epsilon(R^2 + S^2 - T^2)^{1/2}, \qquad Q = T, \qquad \epsilon = \pm 1. \tag{4.7}$$

Throughout the remainder of this section, and also in §5, we shall be concerned with the case in which R and S are not both zero. Since R and S are real functions, this condition is equivalent to $R^2+S^2\neq 0$, or $P^2+Q^2\neq 0$, or $a^*\neq 0$. Then we may solve equations (4.6) for $\sin\phi$ and $\cos\phi$:

$$\sin\phi = \frac{RP - SQ}{P^2 + Q^2}, \qquad \cos\phi = \frac{-SP - RQ}{P^2 + Q^2}.$$

If we substitute the values for P and Q from (4.7), we obtain

$$\begin{aligned} \sin\phi &= \frac{-ST + \epsilon R(R^2 + S^2 - T^2)^{1/2}}{R^2 + S^2}, \\ \cos\phi &= \frac{-RT - \epsilon S(R^2 + S^2 - T^2)^{1/2}}{R^2 + S^2}; \end{aligned} \tag{4.8}$$

these expressions could be obtained more directly by solving equation (4.2) for $\sin\phi$ and $\cos\phi$. When the expressions (4.8) are substituted in (3.1) with $A=M$, we obtain the result stated in the following theorem.

THEOREM 4.2. *If R and S are not both zero, the second fundamental tensor satisfies the identity*

$$\Omega_{\alpha\beta} = Ua_{\alpha\beta} + \epsilon V a^*_{\alpha\beta} + Wg_{\alpha\beta}, \qquad \epsilon = \pm 1, \tag{4.9}$$

where

$$\begin{aligned} a_{\alpha\beta} &= -M_{,\alpha\beta} + (1/N)(M_{,\alpha}N_{,\beta} + M_{,\beta}N_{,\alpha}), \\ U = \frac{-2T}{R^2 + S^2}, \quad V &= \frac{2(R^2 + S^2 - T^2)^{1/2}}{R^2 + S^2}, \quad W = M - \frac{U}{2}g^{\alpha\beta}a_{\alpha\beta}, \end{aligned} \tag{4.10}$$

and R, S and T have the values given in (4.3).

We may use the coordinate system of Theorem 3.1, with $A=M$, and equations (3.2) for the components of $\Omega_{\alpha\beta}$. The theorem then follows readily with the aid of (4.8).

When we substitute the expressions (4.9) in the equations (2.2) of Codazzi, we obtain two third order differential equations on the mean curvature which may be written

$$\epsilon^{\beta\gamma}(Ua_{\alpha\beta} + \epsilon V a^*_{\alpha\beta} + Wg_{\alpha\beta})_{,\gamma} = 0, \tag{4.11}$$

where ϵ has the same value as in (4.9). These are thus necessary conditions on the mean curvature in the case $R^2+S^2\neq 0$.

5. **Sufficiency of the conditions** (4.11). In this section the problem of immersing a given intrinsic surface will be reduced, in the general case, to finding a function $M(u^\alpha)$ satisfying equations (4.11). Exceptional cases will be treated in §6.

THEOREM 5.1. *If $M(u^\alpha)$ is any real solution of equations* (4.11), *then there exists an immersed surface, determined uniquely to within rigid motions in space, for which $2M(u^\alpha)$ is the mean curvature and $g_{\alpha\beta}$ the fundamental tensor.*

Proof. It is readily verified that since (4.11) is satisfied, $\Omega_{\alpha\beta}$ as defined by (4.9) satisfies the equations of Gauss and Codazzi. Thus the existence of the surface follows from classical theory, and its second fundamental tensor is given by (4.9). Moreover, $g^{\alpha\beta}\Omega_{\alpha\beta}=2M$, so that $2M(u^\alpha)$ is actually the mean curvature.

The ambiguity of sign in the expression for $\Omega_{\alpha\beta}$ in (4.9) has the following consequences. If $M(u^\alpha)$ is a solution of (4.11) with $\epsilon=1$, then equations (4.9) with $\epsilon=1$ yield a tensor $\Omega_{\alpha\beta}$ satisfying the equations of Gauss and Codazzi; a similar statement holds for $\epsilon=-1$. If it happens, however, that $M(u^\alpha)$ satisfies the pair of equations

$$\epsilon^{\beta\gamma}(V a^*_{\alpha\beta})_{,\gamma} = 0 \tag{5.1}$$

in addition to (4.11), then the expression (4.9) with ϵ equal to either 1 or -1 will satisfy the equations of Gauss and Codazzi. In this case, from a single function $M(u^\alpha)$ we would be led to two surfaces, intrinsically identical, but having different immersions [3]. We remark that equations (4.11) together with (5.1) would lead to certain necessary conditions on the intrinsic geometry of the surface; these will not be investigated here.

6. **The case** $R=S=0$. In this case it follows from (4.2) that $T=0$; and from the identities (4.6) we have $P=Q=a^*=0$.

THEOREM 6.1. *The condition $P=-\Omega^{*\alpha\beta}a_{\alpha\beta}=0$ is necessary and sufficient that the surface be isometric, that is, that the lines of curvature form an isometric system.*

The proof consists in taking the lines of curvature as parametric, and recalling that the condition that the parametric lines be isometric [4] is $\partial^2 \log (g_{11}/g_{22})/\partial u^1 \partial u^2 = 0$. This condition may be reduced to $P=0$ by making use of the Codazzi equations together with the relations

$$\Omega_{11} = g_{11}(M+N), \qquad \Omega_{22} = g_{22}(M-N).$$

Since $a^* = 0$, from (2.3) we have $a_{11}/g_{11} = a_{12}/g_{12} = a_{22}/g_{22}$. If we write these equations in the form $a_{\alpha\beta} = \rho g_{\alpha\beta}$, and multiply through by $g^{\alpha\beta}$, we obtain $\rho = (1/2)g^{\alpha\beta}a_{\alpha\beta}$. But since $T = 0$, from (4.4) we have

$$g^{\alpha\beta}a_{\alpha\beta} = -(1/2MN^2)(N^2\Delta_2 K + \Delta_1 K + 4N^4 K).$$

Thus

$$a_{\alpha\beta} = -(g_{\alpha\beta}/4MN^2)(N^2\Delta_2 K + \Delta_1 K + 4N^4 K),$$

or

$$\begin{aligned} M_{,\alpha\beta} = {} & (1/N)(M_{,\alpha}N_{,\beta} + M_{,\beta}N_{,\alpha}) \\ & + (g_{\alpha\beta}/4MN^2)(N^2\Delta_2 K + \Delta_1 K + 4N^4 K). \end{aligned} \tag{6.1}$$

Conversely, it is readily shown that equations (6.1) imply $R = 0$, $S = 0$ and $T = 0$.

Equations (6.1) lead to certain necessary conditions on the intrinsic geometry, which we may obtain in the following way. Let us write (6.1) in the form

$$M_{,\alpha\beta} = f_{\alpha\beta}(M_{,\alpha}, M, g_{\alpha\beta}), \tag{6.2}$$

where $f_{\alpha\beta}$ involves the derivatives of $g_{\alpha\beta}$; this will be understood also in the functions F_α and Φ below.

To obtain integrability conditions on (6.2), we make use of the identity

$$\varepsilon^{\beta\gamma}M_{,\alpha\beta\gamma} = -K\varepsilon_{\cdot\alpha}^{\beta}M_{,\beta}. \tag{6.3}$$

This is derived from the identity [5]

$$\lambda_{\alpha,\beta\gamma} - \lambda_{\alpha,\gamma\beta} = \lambda_\delta g^{\delta\mu}R_{\mu\alpha\beta\gamma},$$

in view of the fact that for a two dimensional surface all the components of $R_{\lambda\alpha\beta\gamma}$ vanish except $R_{1212} = R_{2121} = -R_{1221} = -R_{2112} = gK$.

When we differentiate covariantly the expressions (6.2) for $M_{,\alpha\beta}$, substitute in (6.3), eliminate $M_{,\alpha\beta}$ and solve for $M_{,\alpha}$, we get an equation of the form

$$M_{,\alpha} = F_\alpha(M, g_{\alpha\beta}). \tag{6.4}$$

We now develop the integrability condition of (6.4), eliminate $M_{,\alpha}$ and solve for M, obtaining

$$M = \Phi(g_{\alpha\beta}). \tag{6.5}$$

Thus we may obtain a set of necessary conditions on $g_{\alpha\beta}$ by substituting M from (6.5) in (6.2):

$$\Phi_{,\alpha\beta} = f_{\alpha\beta}(\Phi_{,\alpha}, \Phi, g_{\alpha\beta}). \tag{6.6}$$

Conversely, if $g_{\alpha\beta}(u^\gamma)$ is such that (6.6) are identically satisfied, then M defined by (6.5) will be a solution of (6.2).

Certain details have been obtained for the case (of greatest generality) when $\Delta_1(\Delta_1 K) \neq 0$, $\Theta(K, \Delta_1 K) \neq 0$. In this case it turns out that (6.4) can be written in the form

$$M_{,\alpha} = aK_{,\alpha} + b(\Delta_1 K)_{,\alpha}$$

where b satisfies a cubic equation $b^3+Ab^2+Bb+C=0$, and a is given by $a=(Db+E)/(Fb+G)$. The coefficients A, B, C, D, E, F and G are polynomials in M with coefficients formed from the differential parameters of K. The author is not able to give any indication of the form of (6.5).

Suppose that (6.6) are satisfied, and that $M(u^\alpha)$ is a solution of (6.1), such that $\Delta_1 M \neq 0$. Then equations (4.1):

$$\phi_{,\alpha} = \frac{M_{,\gamma}}{N}\,(\varepsilon^{\gamma}_{\cdot\alpha}\cos\phi - \delta^{\gamma}_{\alpha}\sin\phi) - \frac{2\varepsilon^{\beta\gamma}M_{,\beta}M_{,\gamma\alpha}}{\Delta_1 M} - \varepsilon^{\gamma}_{\cdot\alpha}\frac{N_{,\gamma}}{N}$$

are completely integrable. Thus we have an infinity of solutions $\phi(u^\alpha, a)$, where a is an arbitrary parameter. Consider the functions $\Omega_{\alpha\beta}(u^\gamma, a)$ defined by (3.1) with $A=M$:

$$\Omega_{\alpha\beta} = \cos\phi(b_{\alpha\beta} - Ng_{\alpha\beta}) + \sin\phi b^*_{\alpha\beta} + Mg_{\alpha\beta}, \quad b_{\alpha\beta} = 2NM_{,\alpha}M_{,\beta}/\Delta_1 M.$$

It is easily verified that the tensor $\Omega_{\alpha\beta}$ so defined satisfies the equation of Gauss; it also satisfies the equations of Codazzi, since ϕ satisfies (4.1). Moreover, it may be shown, as in Theorem 5.1, that $2M(u^\gamma, a)$ is actually $g^{\alpha\beta}\Omega_{\alpha\beta}(u^\gamma, a)$. Hence we have

THEOREM 6.2. *If $M(u^\alpha)$ is any solution of equations* (6.1) *such that $\Delta_1 M \neq 0$, then there exist* [6] ∞^1 *surfaces with the intrinsic geometry $g_{\alpha\beta}(u^\gamma)$ and the mean curvature $2M(u^\alpha)$.*

The case $\Delta_1 M = 0$, or $M = \text{const.}$, automatically satisfies $R^2+S^2=0$, as is evident from (4.3); but the above discussion is inapplicable, since equations (4.1) are only valid for $\Delta_1 M \neq 0$. We first consider the case $\Delta_1 M = 0$, $\Delta_1 K \neq 0$. Then we may put $A=K$ in equations (3.1) and (3.4). Remembering that $N^2=M^2-K$, and $M_{,\alpha}=0$, we get

$$(6.7)\qquad \Omega_{\alpha\beta} = \cos\phi(b_{\alpha\beta} - Ng_{\alpha\beta}) + \sin\phi b^*_{\alpha\beta} + Mg_{\alpha\beta}, \quad b_{\alpha\beta} = 2NK_{,\alpha}K_{,\beta}/\Delta_1 K,$$

and

$$(6.8)\qquad \phi_{,\alpha} = \frac{-2\varepsilon^{\beta\gamma}K_{,\beta}K_{,\gamma\alpha}}{\Delta_1 K} + \frac{\varepsilon^{\gamma}_{\cdot\alpha}K_{,\gamma}}{2N^2}.$$

If, as in Theorem 4.1, we develop the integrability conditions of (6.8), we obtain the single equation

$$N^2\Delta_2 K + \Delta_1 K + 4N^4 K = 0.$$

The method of proof is identical with that used in Theorem 4.1, except that the curves $K = \text{const.}$ and their orthogonal trajectories are taken to be parametric.

We can eliminate M entirely from this equation by first writing it in the form

$$4KM^4 + M^2(\Delta_2 K - 8K^2) + 4K^3 - K\Delta_2 K + \Delta_1 K = 0. \tag{6.9}$$

When we solve for M^2 and substitute in $\Delta_1(M^2) = 0$, we get

$$\Delta_1 k = 0, \qquad k = \frac{8K^2 - \Delta_2 K \pm (\Delta_2^2 K - 16K\Delta_1 K)^{1/2}}{K}. \tag{6.10}$$

In the particular case $M = 0$, (6.9) reduces to

$$K\Delta_2 K - \Delta_1 K - 4K^3 = 0 \tag{6.11}$$

which is thus a necessary condition on the intrinsic geometry of a minimal surface [7].

Conversely, suppose that we are given an intrinsic surface $g_{\alpha\beta}(u^\gamma)$ satisfying (6.10) with either the + or − sign, say the + sign for definiteness. Then the expression k of (6.10) is a constant. If $k \geqq 0$, then (6.9) is satisfied with $M = (1/2)k^{1/2}$. Hence equations (6.8) are completely integrable, and we have

$$\phi = \int \left[\frac{-2\varepsilon^{\beta\gamma} K_{,\beta} K_{,\gamma\alpha}}{\Delta_1 K} + \frac{\varepsilon_{\cdot\alpha}^{\gamma} K_{,\gamma}}{2N^2} \right] du^\alpha + C.$$

There are thus ∞^1 values of ϕ differing by an additive constant. As in the remark preceding Theorem 6.2 we may conclude that $\Omega_{\alpha\beta}(u^\gamma, C)$ as defined by (6.7) satisfies the equations of Gauss and Codazzi, and that $2M = k^{1/2}$ is the corresponding mean curvature. Hence we have

THEOREM 6.3. *If $g_{\alpha\beta}(u^\gamma)$ satisfies $\Delta_1 k = 0$ where*

$$k = \frac{8K^2 - \Delta_2 K \pm (\Delta_2^2 K - 16K\Delta_1 K)^{1/2}}{K},$$

and if $k \geqq 0$, then there are ∞^1 surfaces Σ with the intrinsic geometry S and the constant mean curvature $k^{1/2}$.

It has already been pointed out that if $\Delta_1 M = \Delta_1 K = 0$, the surface reduces to a plane, a sphere or a circular cylinder.

PART II

7. **The edge of regression: definition.** In view of the significance of the mean curvature in the immersion theory of surfaces, it becomes natural to consider what kind of singularities the function K_m may exhibit, and what corresponding singularities will arise on the immersed surface. We shall now consider a certain curve E on a surface along which K_m is infinite, and which is the analogue for a general surface of the edge of regression of a developable surface. It will be referred to as the *edge of regression*; this name is appropriate not only because of the analogy with a developable surface, but also because it has been used to designate the envelope of a congruence of curves in 3-space [8].

In order to provide an adequate foundation for the study of the edge of regression, both local and in the large, we shall first define an immersed surface with boundary, and then state conditions in order that the boundary may constitute an edge of regression. For convenience, we shall speak in terms of analytic functions; but all the properties obtained would continue to hold if the functions involved possessed only a finite number of derivatives.

The region of the Euclidean plane defined by $|u^\alpha| < \epsilon$ will be called a square; the region defined by $|u^1| < \epsilon$, $0 \leqq u^2 < \epsilon$, will be called a half square, and the segment $u^2 = 0$, $|u^1| < \epsilon$ will be called its bounding edge. The homeomorphic image of a square is of course a region; we shall refer to the homeomorphic image of a half square as a half region, and the image of its bounding edge will be called the bounding edge of the half region.

Consider a two-dimensional manifold together with its boundary. The totality of interior points will be denoted by A, and the totality of boundary points by B. $A+B$ shall be coverable by a finite number of regions R_i, $i=1, 2, \cdots$, and half regions $\overline{R}_j$, $j=1, 2, \cdots$, each homeomorphic to a square S_i or a half square $\overline{S}_j$ of the Euclidean plane. The regions R_i lie on A and the half regions $\overline{R}_j$ have their interior part on A and their bounding edge on B. The coordinates u^α, $\alpha=1, 2$, of a point with respect to a particular R_i or $\overline{R}_j$ to which it belongs are defined to be the Cartesian coordinates of its image under the homemorphism. If a point belongs to more than one of the sets R_i or $\overline{R}_j$, the transformation from one coordinate system to the other will be assumed to be analytic. To each point of $A+B$ there shall correspond an ordered triad of real numbers x^i, $i=1, 2, 3$, such that the sum of the squares of the Jacobians of the x's with respect to the u's is different from zero.

Let Σ be an immersed surface and let the curve E be its bounding edge. E will be assumed to possess a unique tangent at every point. The plane

normal to E through an arbitrary point P of E will cut the surface in a curve C bounded by P. We define as follows a coordinate system $[s, t]$ for a certain half region $\overline{R}$ whose bounding edge lies on E: s is the arc length along E from a fixed point O to P, and t is the arc length along C measured from P. The curve E will be called an *analytic edge of regression* if it satisfies the following three conditions:

CONDITION I. *The mean curvature is infinite and the total curvature finite along E.*

CONDITION II. *The space coordinates x^i of the curves $t=$ const. are functions of s analytic in $\overline{R}$.*

CONDITION III. *The curves C: $s=$ const. have an algebraic singularity at $t=0$; that is, within $\overline{R}$ the space coordinates x^i of C are analytic in $t^{1/\alpha}$, where α is an integer called the order of the singularity. α is assumed to be the smallest integer for which this condition is satisfied. E is not a straight line.*

8. **Analytic structure of the edge of regression.** We may now write the equations of the surface in terms of the parametrization $[s, t]$ defined in §7. Let $x^i=x^i(s)$ be the equation of E, and let $\xi_1^i(s)$, $\xi_2^i(s)$ and $\xi_3^i(s)$ denote, respectively, the unit tangent, principal normal and binormal vectors of E. Then the equations of the surface may be written in the form

$$X^i = x^i + a(s, t)\xi_2^i + b(s, t)\xi_3^i, \tag{8.1}$$

where $a(s, t)$ and $b(s, t)$ must, according to Condition II, be analytic in s; while according to Condition III, they must be analytic in $t^{1/\alpha}$. Hence

$$a = \sum_1^\infty a_\mu t^{\mu/\alpha}, \qquad b = \sum_1^\infty b_\mu t^{\mu/\alpha},$$

where a_μ and b_μ are analytic in s.

From the definition of the parameter t, it follows that $g_{22}=1$, where $g_{\alpha\beta}$ denotes the fundamental tensor in the $[s, t]$ system. We may use this fact to prove that the integer α defined in Condition II (the order of the singularity) is equal to 2. We have

$$\frac{\partial X^i}{\partial t} = \frac{1}{\alpha}\left[\sum_1^\infty \mu a_\mu t^{\mu/\alpha-1}\right]\xi_2^i + \frac{1}{\alpha}\left[\sum_1^\infty \mu b_\mu t^{\mu/\alpha-1}\right]\xi_3^i.$$

Hence

$$g_{22} = \frac{1}{\alpha^2}\sum_{\mu=1}^\infty \sum_{j=1}^{\mu+2\alpha-1} j(\mu + 2\alpha - j)\left[a_j a_{\mu+2\alpha-j} + b_j b_{\mu+2\alpha-j}\right]t^{\mu/\alpha}.$$

But since $g_{22}=1$, we have, for $\mu \not\equiv 0$,

$$(8.2.\mu) \qquad \sum_{j=1}^{\mu+2\alpha-1} j(\mu+2\alpha-j)[a_j a_{\mu+2\alpha-j} + b_j b_{\mu+2\alpha-j}] = 0.$$

We first show that $\alpha \not\equiv 1$. If α were 1, the coordinates of the surface would be analytic in the parameters $[s, t]$, and the mean curvature would be finite, contrary to Condition I.

Secondly, α cannot be odd. For suppose α were odd and not 1. We shall prove by induction that

$$(8.3) \qquad a_\mu = b_\mu = 0 \qquad \text{for } \mu \not\equiv 0 \pmod{\alpha}.$$

Equation $(8.2.2-2\alpha)$ yields $a_1^2+b_1^2=0$ or

$$(8.4) \qquad a_1 = b_1 = 0.$$

Suppose we have shown that $a_\nu=b_\nu=0$ for $\nu \not\equiv 0 \pmod{\alpha}$, $\nu<\mu$. Consider equation $(8.2.2\mu-2\alpha)$, where $\mu \not\equiv 0 \pmod{\alpha}$. Every one of the products $a_\lambda a_{2\mu-\lambda}$, $b_\lambda b_{2\mu-\lambda}$ occurring in this equation will involve a term a_ν or b_ν for which $\nu \not\equiv 0 \pmod{\alpha}$ and $\nu<\mu$, except the terms $a_\mu^2+b_\mu^2$. Hence all the product terms will vanish except $a_\mu^2+b_\mu^2$, and equation $(8.2.2\mu-2\alpha)$ will yield $a_\mu^2+b_\mu^2=0$, since $2\mu-2\alpha \not\equiv 0$. Hence $a_\mu=b_\mu=0$; and equation (8.3) is established in general, since we have shown independently that $a_1=b_1=0$. But (8.3) implies that all the terms in the expansions of a and b involving fractional powers of t drop out. Thus a and b would be analytic in t, which we have already shown to be inconsistent with Condition I.

Finally, suppose α is even. Then an argument exactly analogous to that of the previous paragraph would enable us to prove that $a_\mu=b_\mu=0$ for $\mu \not\equiv 0 \pmod{\alpha/2}$. Hence a and b would be analytic in $t^{1/2}$; and from the minimal property of α, as stated in Condition III, we must have $\alpha=2$.

Let us put $\theta=t^{1/2}$. Then we have

$$(8.5) \qquad a = \sum_{\mu=1}^{\infty} a_\mu \theta^\mu, \qquad b = \sum_{\mu=1}^{\infty} b_\mu \theta^\mu.$$

The remainder of this section will be occupied with showing that

$$(8.6) \qquad a_1 = b_1 = 0, \qquad a_2 = \epsilon = \pm 1, \qquad b_2 = a_3 = 0, \qquad b_3 \neq 0.$$

The relations $a_1=b_1=0$ have already been obtained in (8.4). We shall use the notation $A=o(\theta^n)$ to mean that $\lim_{\theta \to 0} A/\theta^n = k$ where k is a function of s alone. Then we may write (8.1) in the form

$$(8.7) \qquad X^i = x^i + \xi_2^i[a_2\theta^2 + a_3\theta^3 + o(\theta^4)] + \xi_3^i[b_2\theta^2 + b_3\theta^3 + o(\theta^4)].$$

From these we obtain, using the Frenet formulas, and denoting by ρ and τ, respectively, the radii of curvature and torsion of E,

$$
\begin{aligned}
&g_{11} = 1 + o(\theta^2), \qquad g_{12} = o(\theta), \qquad g_{22} = 1, \qquad g = 1 + o(\theta^2),\\
&\Omega_{11} = \frac{-b_2}{\rho} + o(\theta),\\
&\Omega_{12} = -b_2\left(\frac{\partial a_2}{\partial s} - \frac{b_2}{\tau}\right) + a_2\left(\frac{\partial b_2}{\partial s} + \frac{a_2}{\tau}\right) + o(\theta),\\
&\Omega_{22} = \frac{3}{4\theta}\left[-a_3b_2 + a_2b_3 + o(\theta)\right].
\end{aligned}
\tag{8.8}
$$

The mean curvature and the total curvature are thus found to be

$$K_m = (3/4\theta)\left[-a_3b_2 + a_2b_3 + o(\theta)\right], \tag{8.9}$$

$$K = -(3b_2/4\rho\theta)\left[-a_3b_2 + a_2b_3 + o(\theta)\right]. \tag{8.10}$$

Since, according to Condition III, E is not a straight line, $1/\rho \not\equiv 0$. Let us consider the neighborhood of a point at which $1/\rho \not\equiv 0$. By Condition I, K_m is infinite for $\theta = 0$, and hence from (8.9)

$$a_3b_2 - a_2b_3 \not\equiv 0. \tag{8.11}$$

Again by Condition I, K is finite for $\theta = 0$, so that from (8.10) we must have $b_2 = 0$. This fact, together with (8.11), implies that $a_2 \not\equiv 0$, $b_3 \not\equiv 0$. Equation (8.2.0) yields $a_2^2 + b_2^2 = 1$, whence $a_2 = \epsilon = \pm 1$. Finally, from equation (8.2.1) we have

$$10(a_2a_3 + b_2b_3) = 0,$$

so that $a_3 = 0$. This completes the proof of equations (8.6).

In the Cartesian coordinates $[y^1, y^2]$, consider the curve $y^1 = a$, $y^2 = b$. In view of (8.6), these equations become

$$y^1 = \sum_2^\infty a_\mu\theta^\mu, \qquad y^2 = \sum_3^\infty b_\mu\theta^\mu. \tag{8.12}$$

These are the equations of a curve with a cusp at the origin, of which one branch is the curve C, and whose cuspidal tangent coincides with ξ_2^1. Let us choose the sign of a_2 in such a way that C corresponds to $\theta > 0$, and let us denote by C' the branch of the curve corresponding to $\theta < 0$. As the point P traverses E, the curve C will generate the surface Σ, and the curve C' will generate a second surface Σ' meeting Σ cuspidally along E. The surfaces Σ

and Σ', corresponding respectively to $\theta>0$ and $\theta<0$, will be referred to as the two sheets of the edge of regression.

Conversely, if we are given a plane curve whose equations in the $[y^1, y^2]$ plane are of the form (8.12), with $a_2\neq 0$, $a_3=0$, $b_3\neq 0$, with a^i and b^i analytic in s, and with θ^2 equal to the arc length of the curve, measured from the cusp, then the surface

$$X^i = x^i + \xi_2^i \sum_2^\infty a_\mu \theta^\mu + \xi_3^i \sum_3^\infty b_\mu \theta^\mu,$$

which is generated by displacing the curve along the analytic curve E: $X^i=x^i$ in such a way that the point $\theta=0$ lies on E and the cuspidal tangent at $\theta=0$ coincides with the principal normal of E, obviously has E as an analytic edge of regression.

9. **Local geometrical properties of the edge of regression.** This section will be devoted to deriving certain geometrical properties of the edge of regression. Consider a surface defined, in the neighborhood of its edge of regression E, by

$$X^i = x^i + \xi_2^i \sum_2^\infty a_\mu \theta^\mu + \xi_3^i \sum_3^\infty b_\mu \theta^\mu,$$

where X^i is analytic in the half region $\overline{R}$: $|s|<r_1$, $0\leqq\theta^2<r_2$. The geodesic curvature of E may be defined by

$$\frac{1}{\rho_g} = \frac{-1}{g^{1/2}} \frac{\partial g_{11}^{1/2}}{\partial t}. \tag{9.1}$$

The half region $\overline{R}$ is said to be *convex* (concave) *with respect to its bounding edge* E at a point P of E if the geodesic curvature of E is greater than 0 (less than 0) at P.

We may mention in passing that the concept of convexity as defined above is equivalent to the following more intuitive definition, which may be given a rigid analytical foundation: the half region $\overline{R}$ is convex (concave) with respect to E at P if the geodesics orthogonal to E converge (diverge) in $\overline{R}$ near P.

PROPERTY I. *If* $1/\rho$, $1/\rho_g$, $1/\tau$ *and* K *denote, respectively, the curvature, geodesic curvature and torsion of* E, *and the total curvature of the surface along* E, *then*

$$\rho = \epsilon\rho_g, \qquad K + 1/\tau^2 \neq 0, \qquad \epsilon = \pm 1. \tag{9.2}$$

Proof. From (8.7) we have

$$g_{11} = \frac{\partial x^i}{\partial s}\frac{\partial x^i}{\partial s} = 1 - 2\frac{a_2\theta^2}{\rho} + o(\theta^3).$$

Hence, from (9.1),

$$1/\rho_g = a_2/\rho + o(\theta^2). \tag{9.3}$$

But since, from (8.6), $a_2 = \epsilon = \pm 1$, it follows that $\rho = \epsilon\rho_g$ along E.

Combining equations (8.8) and (8.6), we have

$$\Omega_{11} = \frac{-3b_3}{2\rho}\theta + o(\theta^2), \quad \Omega_{12} = \frac{1}{\tau} + o(\theta), \quad \Omega_{22} = \frac{3a_2b_3}{4\theta}[1 + o(\theta)], \tag{9.4}$$

whence $K = -9a_2b_3^2/8\rho - 1/\tau^2 + o(\theta)$, so that, along E,

$$K + \frac{1}{\tau^2} = -\frac{9a_2b_3^2}{8\rho} = -\frac{9b_3^2}{8\rho_g} \neq 0. \tag{9.5}$$

The case $1/\rho = 1/\rho_g = 0$ has already been excluded.

PROPERTY II. *A surface is convex or concave with respect to its edge of regression according as $K + 1/\tau^2 < 0$ or > 0 along the edge.*

Proof. From the definition of convexity, the surface will be convex or concave with respect to its edge of regression according as ρ_g is positive or negative along E; that is, because of (9.5), according as $K + 1/\tau^2 < 0$ or > 0 along E.

PROPERTY III. *The edge of regression is the envelope of both sets of asymptotic lines, which are analytic curves near E. It is likewise enveloped by one set of lines of curvature, and met orthogonally by the other set.*

Proof. The asymptotic lines are given by

$$\Omega_{11}ds + (\Omega_{12} \pm (-gK)^{1/2})dt = 0.$$

We shall consider only the set of asymptotic lines for which the $+$ sign is appropriate; the proof is identical for the other set. Making use of (8.8) and (9.4) we have

$$\left[\frac{3b_3}{2\rho}\theta + o(\theta^2)\right]ds + \left[\frac{1}{\tau} + (-K)^{1/2} + o(\theta)\right]dt = 0,$$

so that

$$\frac{dt}{ds} = \frac{3b_3\theta}{2\rho(1/\tau + (-K)^{1/2})} + o(\theta^2). \tag{9.6}$$

This shows that $dt/ds = 0$ for $\theta = 0$, which implies that the asymptotic lines

are tangent to E. Since the directions of the lines of curvature bisect the angles between the asymptotic lines, it is obvious that one set of lines of curvature will envelope E, and the other set will meet it orthogonally.

Putting $dt=2\theta d\theta$ in (9.6) and inverting, we have

$$\frac{ds}{d\theta}=\frac{-4\rho}{3b_3}(1/\tau+(-K)^{1/2})+o(\theta).$$

Since the right-hand side is analytic in s and θ, it follows that s is analytic in θ. Let σ be the arc length along the asymptotic line. Then, recalling that $g_{22}=1$, we have

$$\left(\frac{d\sigma}{d\theta}\right)^2=g_{11}\left(\frac{ds}{d\theta}\right)^2+2g_{12}\left(\frac{ds}{d\theta}\right)+1, \tag{9.7}$$

which shows that $d\sigma/d\theta$ is analytic in θ, and hence that σ is analytic in θ. But from (9.7), $d\sigma/d\theta$ is obviously not 0 for $\theta=0$, and we conclude that θ is analytic in σ, and that the same is true for s. Finally, substituting these analytic functions in the equations (8.1) of the surface, we obtain the coordinates of a point on the asymptotic line as analytic functions of its arc length.

We may also show that the asymptotic lines cross over from Σ to Σ' as they envelope E. This follows from the fact that, according to equation (9.7), $d\sigma/d\theta\neq 0$ at E. Hence, near E a change in the sign of σ leads to a change in the sign of θ, and thus to a change from Σ to Σ', or vice versa.

10. **The index of a region.** Consider an immersed surface whose only singularity is an edge of regression, consisting of one or several pieces. We wish to study the topology of the region bounded by such a singular locus. For this purpose, §10 will develop the properties of an index of a region, intimately connected with its Euler characteristic. In §11 the index of a region bounded by an edge of regression will be shown to depend upon the indices of the umbilics in that region.

Indices of the type we are considering, and in particular the index of an umbilic, have been studied by Hamburger [9], Blaschke [10], and Franklin [11]. In their treatment of umbilics, these authors depend on the work of Darboux [12] and Gullstrand [13], who investigated the behaviour of the lines of curvature in the neighborhood of an umbilic. The present treatment is quite independent of these earlier discussions, and makes use of the invariants of the surface at an umbilic, rather than the descriptive geometry of the lines of curvature near an umbilic.

We shall now proceed to define and discuss an index j_R for a region R of an orientable surface.

Let R be a region of an orientable surface Σ, such that the boundary C of R contains no umbilics and has a unique tangent at all except a finite number of points P_i, $i=1, 2, \cdots$, at each of which it has a corner. Let C be oriented in such a way that the positive tangent vector T of C, the normal N of C directed toward R, and the normal of Σ form a positive triad of directions. Consider the vector T as it traverses C in the positive sense; let θ_i be the angle through which T must rotate in passing from one to the other of the two segments of C which meet at the angle P (the positive sense of rotation is from T to N). Let P be any point of C and let V be the unit tangent vector to an arbitrarily chosen line of curvature through P, with an arbitrary orientation. Finally, let θ be the positive angle from V to T. Then we may define an index j_R for the region R by

$$j_R = 1 - \frac{1}{2\pi}\Big[\oint_C d\theta + \sum_i \theta_i\Big].$$

Consider the vector V of the above definition, and let V' be the tangent vector of the other line of curvature through P. It is readily seen that, as P makes a circuit of C, the vectors V and V' cannot be permuted. For if $1/\rho_1$ and $1/\rho_2$ are the principal curvatures corresponding to V and V' respectively, and if $1/\rho_1 > 1/\rho_2$ at the start, then $1/\rho_1 > 1/\rho_2$ over the whole path, since it contains no umbilics. This implies that V and V' will still correspond to the same principal directions after any circuit.

The vector V may, however, return to its original position with sense reversed. An example illustrating this fact is afforded by a circuit enclosing an umbilic of an ellipsoid, for which the lines of curvature in the neighborhood of an umbilic have the form illustrated in Fig. 1. Consider the region R: $ABCD$, enclosed by two lines of curvature, with right angles at B and D. The index is $j_R = 1-(\pi/2\pi) = 1/2$, and it is clear that the vector V returns to its original position with sense reversed.

Since V must be carried into $\pm V$ after one circuit, it follows that $\oint_C d\theta + \sum_i \theta_i$ must be some multiple of π. Hence j_R must be some multiple, positive or negative, of $1/2$. This has the consequence that the index is unchanged when R is deformed continuously in such a way that its boundary never goes through an umbilic.

THEOREM 10.1. *If R is a region which is simply connected and free from umbilics, then $j_R = 0$.*

Proof. Consider any conformal mapping

$$\Sigma \sim E \tag{10.1}$$

which carries the region R, the curve C, and the vectors V and T of Σ, respectively, into $\overline{R}$, $\overline{C}$, $\overline{V}$ and $\overline{T}$ of the Euclidean plane E. Let D be a fixed oriented direction in the plane. Let ϕ and ψ denote, respectively, the positive angle from D to $\overline{V}$ and the positive angle from D to $\overline{T}$. Then $\theta=\psi-\phi$.

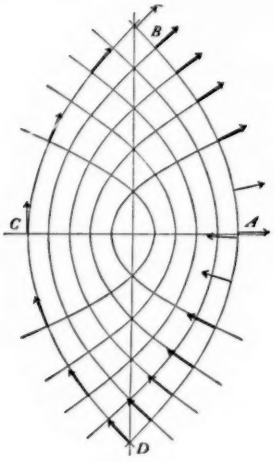


FIG. 1

Since $\overline{R}$ is simply connected, $\overline{C}$ is a simple curve, and the vector $\overline{T}$ will turn through 2π in one circuit, which is the total increment in ψ: $\oint_{\overline{C}} d\psi + \sum_i \theta_i = 2\pi$. Hence

$$j_R = \frac{1}{2\pi} \oint_{\overline{C}} d\phi.$$

Since R is free from umbilics, V is a nonsingular vector field over R, and $\overline{V}$ is a nonsingular vector field over $\overline{R}$. Thus the angle ϕ is a nonsingular scalar function over $\overline{R}$, and the increment of ϕ on traversing the boundary $\overline{C}$ of $\overline{R}$, $\oint_{\overline{C}} d\phi$, equals zero. Hence $j_R=0$.

THEOREM 10.2. *If R is a region which contains p holes and is free from umbilics,*

$$j_R = p = 1 - \chi(R),$$

where $\chi(R)$ is the Euler characteristic of R [14].

Proof. We reduce R to a simply connected region R' by introducing p cuts, one cut connecting each hole to the outer boundary of R. Then R' is

simply connected and free from umbilics, so that $j_{R'}=0$. But in making the cuts, we have introduced $4p$ new angles on the boundary, whose total contribution to $\sum_i \theta_i$ is $2p\pi$; so that the index of R has been reduced by p. Hence $j_R = j_{R'} + p = p$.

THEOREM 10.3. *Let R be a simply connected region covered by a system of isometric coordinates, so that*

$$g_{11} = g_{22} = \lambda, \qquad g_{12} = 0$$

and such that the boundary C of R is free from umbilics and corners. Then

$$j_R = \frac{1}{4\pi}\oint_C d\left(\arcsin \frac{\Omega_{12}}{\lambda N}\right).$$

Proof. Let α be the angle from the line $u^1=\text{const.}$ to the vector V previously defined. Then under the conformal mapping (10.1), the direction $du^1=0$ is carried into a nonsingular vector field over $\bar{R}$. Hence $\phi-\alpha$, the angle from D to $du^1=0$, is a nonsingular scalar function over $\bar{R}$, so that $\oint_{\bar{C}} d(\phi-\alpha)=0$. Thus

$$j_R = \frac{1}{2\pi}\oint_{\bar{C}} d\phi = \frac{1}{2\pi}\oint_{\bar{C}} d\alpha.$$

Let us put $a=\Omega_{12}$, $b=(1/2)(\Omega_{11}-\Omega_{22})$. Then the lines of curvature are given by

$$-a(du^1)^2 + 2b du^1 du^2 + a(du^2)^2 = 0,$$

from which we find $\sin 2\alpha = a/\lambda N = \Omega_{12}/\lambda N$, and the theorem is proved.

11. **Umbilics. Region bounded by the edge of regression.** In this section, Theorem 10.3 will be used to evaluate the index of a region containing an umbilic of the simplest type, in terms of a certain invariant evaluated at the umbilic. The method is perfectly general, however, and could be extended to umbilics of higher order; the present results are intended merely as illustrative. It is then possible to evaluate the Euler characteristic of a region bounded by an edge of regression. We shall assume that the coordinates of the surface in the neighborhood of the umbilic are functions of class C^{iv} in the parameters. We shall also assume that the parametrization $[u, v]$ is such that the positive tangent to $v=\text{const.}$, the positive tangent to $u=\text{const.}$, and the surface normal form a positive triad.

We shall deal only with the simplest type of umbilic, namely, those for which

$$\Gamma = g^{\alpha\beta}\varepsilon^{\gamma\delta}\varepsilon^{\lambda\mu}\Omega_{\alpha\gamma\lambda}\Omega_{\beta\delta\mu} \neq 0, \tag{11.1}$$

where $\Omega_{\alpha\beta\gamma}=\Omega_{\alpha\beta,\gamma}$; we have suppressed the comma since $\Omega_{\alpha\beta,\gamma}$ is symmetric in all three indices.

THEOREM 11.1. *At an umbilic the invariant* Γ *satisfies the identity*

$$\Gamma = \Delta_2 K - \frac{K_m}{2}\Delta_2 K_m,$$

where K *and* K_m *are, respectively, the total and mean curvatures.*

Proof. In an isometric coordinate system we find that both sides of the identity reduce to

$$\Gamma = (2/\lambda^3)[\Omega_{111}\Omega_{122} + \Omega_{112}\Omega_{222} - \Omega_{112}^2 - \Omega_{122}^2]. \tag{11.2}$$

If the simply connected region R contains a single umbilic, then j_R is called the *index of the umbilic*. We shall use the notation sign (x), or sometimes sign x, to denote a quantity which is 1, -1 or 0 according as x is positive, negative or zero.

THEOREM 11.2. *If* U *is an umbilic for which* $\Gamma\neq 0$, *then its index* j_U *is given by*

$$j_U = (1/2)\ \text{sign}\ (\Gamma)$$

where Γ *is given by* (11.1).

Proof. Let R be a simply connected region surrounding U, covered by an isometric coordinate system $[u, v]$ of which the umbilic is the origin, and bounded by the curve k:

$$au + bv = r\sin\beta, \qquad cu + dv = r\cos\beta, \tag{11.3}$$

where r is constant and

$$a = \Omega_{112}, \quad b = \Omega_{122}, \quad c = (1/2)(\Omega_{111} - \Omega_{122}), \quad d = (1/2)(\Omega_{112} - \Omega_{222}),$$

all evaluated at the origin. We shall assume that at the origin $g_{11}=g_{22}=\lambda=1$. Making use of (11.2), it is readily verified that at the origin, $bc-ad=\Gamma/4$.

Using Taylor's series, we find the following evaluations for Ω_{12} and N at points of k:

$$\Omega_{12} = r\sin\beta + o(r^2), \qquad N = r + o(r^2).$$

Hence, on k,

$$\sin 2\alpha = \Omega_{12}/\lambda N = \sin\beta + o(r), \qquad 2\alpha = \beta + o(r).$$

Consider in the $[u, v]$ plane the conic H defined by (11.3). Since $\Gamma\neq 0$, it is readily shown that H is a nondegenerate ellipse. The curve k will like-

wise be a nondegenerate closed curve; and as the curve k is traversed in the positive sense, the curve H will also be traversed in the positive sense. That is, the angle γ defined by $\gamma = \arctan v/u$ will run from 0 to 2π.

The relationship between β and γ is given, from (11.3), by

$$\tan\beta = \frac{a + b\tan\gamma}{c + d\tan\gamma}. \tag{11.4}$$

This shows that as $\tan\gamma$ runs from $-\infty$ to ∞, $\tan\beta$ covers the same range in the positive or negative sense. That is, as γ increases by 2π, β increases by $\pm 2\pi$. From (11.4) we obtain

$$\frac{d\beta}{d\gamma} = \frac{\Gamma\cos^2\beta\sec^2\gamma}{4(c + d\tan\gamma)^2},$$

which shows that as γ increases by 2π, β increases by 2π sign (Γ). Hence

$$j_U = \frac{1}{4\pi}\oint_k 2d\alpha = \frac{1}{4\pi}\oint_k d\beta + o(r) = (1/2)\text{ sign }(\Gamma) + o(r).$$

But j_U is independent of the size of the curve k, and so $j_U = (1/2)$ sign (Γ).

THEOREM 11.3. *If R is a region entirely bounded by the edge of regression, free from singularities and containing umbilics U_i, $i = 1, 2, \cdots$, with indices j_i, then the Euler characteristic of R is given by $\chi(R) = \sum_i j_i$.*

Proof. The edge of regression cannot contain umbilics, since along E $K_m = \infty$ and K is finite, while at an umbilic $K_m^2 = 4K$. Since the boundary E of R is met orthogonally by one set of lines of curvature, the angle θ appearing in the definition of j_R is constantly equal to 0, $\pi/2$, π or $3\pi/2$. Hence by definition, $j_R = 1$.

Suppose that there are q umbilics, and that R has p holes. Let R_i, $i = 1, 2, \cdots, q$, be a set of nonoverlapping simply connected regions, each with a boundary C_i which contains neither umbilics nor corners, and such that $U_i \subset R_i$. Then

$$j_{R_i} = j_i = 1 - \frac{1}{2\pi}\oint_{C_i} d\theta.$$

On the other hand, the region $R' = R - \sum_i R_i$ is free from umbilics and has $p+q$ holes, and its boundary is $E - \sum_i C_i$. Hence

$$\begin{aligned} j_{R'} = p + q &= 1 - \left[\oint_E d\theta - \sum_i \oint_{C_i} d\theta\right] = 1 - \sum_i (1 - j_{R_i}) \\ &= 1 + q - \sum_i j_i, \end{aligned}$$

so that

$$\chi(R) = 1 - p = \sum_i j_i.$$

In the case when the umbilics U_i are all such that $\Gamma \neq 0$, we may combine Theorems 11.2 and 11.3 to obtain

THEOREM 11.4. *If R is a region entirely bounded by the edge of regression, free from singularities, and containing umbilics U_i, $i=1, 2, \cdots$, such that $\Gamma(U_i) \neq 0$, then the Euler characteristic of R is given by*

$$\chi(R) = \frac{1}{2} \sum_i \operatorname{sign} \Gamma(U_i).$$

It is, of course, always to be understood that E may consist of one or several pieces. As simple illustrations of Theorem 11.3, consider the surfaces of revolution whose meridian curves are shown in Fig. 2. In each case BC

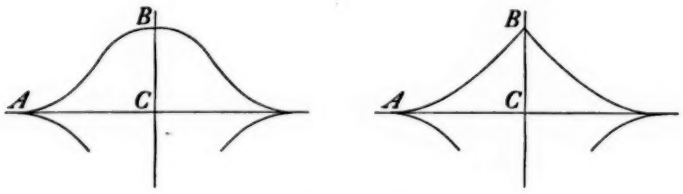


FIG. 2

is the axis of revolution. In the first example, we obtain a simply connected region, free from singularities, and bounded by the edge of regression traced out by the cusp at A. For such a region, $\chi(R) = 1$; hence Theorem 11.3 implies that there must be one or more umbilics. As a matter of fact, B is an umbilic of index $j_U = 1$. However, $\Gamma = 0$ for that umbilic, so that this illustration does not come under Theorem 11.4. In the second example the region is again simply connected, so that $\chi(R) = 1$. Since the surface is of negative curvature, there can be no umbilics. Thus Theorem 11.3 necessitates the existence of a singularity—in this case a conical point at B.

NOTES AND REFERENCES

1. W. C. Graustein, *Applicability with preservation of both curvatures*, Bulletin of the American Mathematical Society, vol. 30 (1924), pp. 19–23; Duke Mathematical Journal, vol. 2 (1936), pp. 177–191.

2. The tensor $\epsilon^{\alpha\beta}$ is defined and discussed in A. Duschek, *Lehrbuch der Differentialgeometrie*, Leipzig, Teubner, 1930, p. 99, formula (19).

3. The conditions (5.1) correspond to the conditions "$z = Q/P$ satisfies the equations (4)" of

Theorem 1 in Graustein's paper in the Duke Mathematical Journal.

4. A. R. Forsyth, *Differential Geometry*, Cambridge, University Press, 1920, p. 84.

5. L. P. Eisenhart, *Riemannian Geometry*, Princeton, University Press, 1926, p. 30, equation (11.14).

6. See Theorem 1 in Graustein's paper in the Duke Mathematical Journal.

7. Equation (6.11) is another form of the condition $\Delta_2 \log(-K) - 4K = 0$, which is due to Ricci. See L. Bianchi, *Vorlesungen über Differentialgeometrie*, translated by M. Lukat, Leipzig, Teubner, 1899, p. 382.

8. Goursat-Hedrick, *A Course in Mathematical Analysis*, vol. 2, part 2, New York, Ginn, 1917, p. 209.

9. H. Hamburger, *Über Kurvenetze mit isolierten singularitäten auf geschlossenen Flächen*, Mathematische Zeitschrift, vol. 19 (1924), pp. 50–56.

10. W. Blaschke, *Über die Geometrie von Laguerre*, IV. *Von den Nabelpunkten einer Eifläche*, Mathematische Zeitschrift, vol. 24 (1926), pp. 617–621.

11. P. Franklin, *Regions of positive and negative curvature on closed surfaces*, Journal of Mathematics and Physics, vol. 13 (1934), pp. 253–260.

12. G. Darboux, *Théorie des Surfaces*, Paris, Gauthier-Villars, 1896, vol. 4, pp. 448–465.

13. A. Gullstrand, *Zur Kenntniss der Kreispunkte*, Acta Mathematica, vol. 29 (1905), pp. 59–100.

14. See, for example, S. Lefschetz, *Topology*, American Mathematical Society Colloquium Publications, vol. 12, New York, 1930, p. 44.

LEHIGH UNIVERSITY,
BETHLEHEM, PA.

ON THE DEGREE OF CONVERGENCE OF SEQUENCES OF RATIONAL FUNCTIONS*

BY

J. L. WALSH

The writer has recently studied† the convergence of certain sequences of rational functions of the complex variable, under the hypothesis that the poles of these functions are prescribed and satisfy certain asymptotic conditions. The rational functions are determined either by interpolation to a given analytic function, or by some extremal property of best approximation to such a function. Degree of convergence and regions of uniform convergence of the sequence of rational functions are then obtained (op. cit.). It is the object of the present paper to go more deeply than hitherto into properties of degree of convergence of sequences of rational functions, to make more precise the previous results, and especially to introduce and study the concept of maximal convergence of a sequence of rational functions with preassigned poles; this is a generalization of the corresponding concept for sequences of polynomials. The analogy between convergence properties of sequences of polynomials and convergence properties of sequences of more general rational functions is strong, but has hitherto not been sufficiently strong to justify the use of the term maximal convergence in the latter case (compare op. cit., p. 258). We show now that maximal convergence is characteristic of various sequences of rational functions determined by interpolation and by extremal properties. The present results would seem to be more or less definitive in form.

1. **Introductory results.** We choose as point of departure the following relatively simple but typical formulation (op. cit., §8.3):

THEOREM 1. *Let R be an annular region bounded by two Jordan curves C_1 and C_2, with C_2 interior to C_1. Let the points $\alpha_{n1}, \alpha_{n2}, \cdots, \alpha_{nn}$ lie exterior to C_1, and let the points $\beta_{n1}, \beta_{n2}, \cdots, \beta_{n,n+1}$ lie on or interior to C_2.‡ Let $\overline{R}$ denote the*

---

* Presented to the Society, December 30, 1938, under the title *Maximal convergence of sequences of rational functions*; received by the editors September 22, 1939.

† *Interpolation and Approximation by Rational Functions in the Complex Domain*, American Mathematical Society Colloquium Publications, vol. 20, New York, 1935. See especially Chapters VIII and IX. Unless otherwise indicated, all references in this paper are to this work, to which the reader should refer also for terminology.

‡ It is a matter of taste whether or not to allow points α_{nk} to lie on C_1 and points β_{nk} to lie on C_2, and whether or not to require that (1) should hold in $\overline{R}$ or on suitably restricted closed sets in $\overline{R}$. There are a variety of allowable choices here. The one we have adopted seems to the writer the most

closure of R. Let the relation

$$(1) \qquad \lim_{n\to\infty} \left| A_n \frac{(z-\beta_{n1})(z-\beta_{n2})\cdots(z-\beta_{n,n+1})}{(z-\alpha_{n1})(z-\alpha_{n2})\cdots(z-\alpha_{nn})} \right|^{1/n} = \Phi(z)$$

hold uniformly on any closed set in $\overline{R}$ containing no point of C_2. Let the function $\Phi(z)$ be continuous in the closed region $\overline{R}$, and take constant values γ_1 and $\gamma_2<\gamma_1$ on C_1 and C_2 respectively. Denote generically by C_γ the curve $\Phi(z)=\gamma$ in $\overline{R}$, $\gamma_2\leqq\gamma\leqq\gamma_1$.*

If the function $f(z)$ is analytic throughout the interior of C_γ but is not analytic throughout the interior of any $C_{\gamma'}$, with $\gamma'>\gamma$, then we have $(\gamma_2<\lambda<\gamma)$

$$(2) \qquad \limsup_{n\to\infty} [\max \mid f(z)-r_n(z)\mid,\ z \text{ on } C_\lambda]^{1/n} \leqq \lambda/\gamma,$$

and we have also the limiting case of (2):

$$(3) \qquad \limsup_{n\to\infty} [\max \mid f(z)-r_n(z)\mid,\ z \text{ on } C_2]^{1/n} \leqq \gamma_2/\gamma,$$

where $r_n(z)$ is the rational function of degree n whose poles lie in the points $\alpha_{n1}, \alpha_{n2}, \cdots, \alpha_{nn}$, and which interpolates to $f(z)$ in the points $\beta_{n1}, \beta_{n2}, \cdots, \beta_{n,n+1}$.

Inequality (3) is a direct consequence of (2), by means of the relation

$$[\max \mid f(z)-r_n(z)\mid,\ z \text{ on } C_2] \leqq [\max \mid f(z)-r_n(z)\mid,\ z \text{ on } C_\lambda],$$

and by allowing λ in (2) to approach γ_2.

The form (1) obviously breaks down whenever a point α_{nk} is infinite, a highly important case which we do not intend to exclude. We therefore use the convention (op. cit., §§8.1, 8.2, 8.5) that in such an expression as the left-hand member of (1) one or more of the points α_{nk} may be infinite; under

convenient in view of the applications. If other choices are made, the conclusions corresponding to Theorem 1 can be read off at once from the present Theorem 1. In later parts of the present paper other choices seem more desirable. There is an obvious asymmetry in Theorem 1 relative to α_{nk} and C_1 on the one hand, and β_{nk} and C_2 on the other hand. This is due to the fact that we assume $f(z)$ analytic on C_2 but not on C_1, and desire to study degree of convergence of $r_n(z)$ to $f(z)$ on C_2. It is then desirable to allow the points β_{nk} to lie on (not merely within) C_2; on occasion the points β_{nk} are to be chosen uniformly distributed on C_2. We need, however, a curve C_1 or $C_\mu, \gamma_1\geqq\mu>\gamma$, in such a relation as (4); it is desirable for explicitness to allow μ to be γ_1, hence desirable to assume (1) valid on C_1 and undesirable to allow points α_{nk} to lie on C_1. That is to say, points β_{nk} are readily and conveniently admitted to $\overline{R}$, but not points α_{nk}.

In Theorem 1 the demands on the location of the α_{nk} and β_{nk} may without change in proof be replaced by the demands that no more than a finite number of the α_{nk} shall lie on or within C_1, and that the β_{nk} shall have no limit point exterior to C_2. But if this new hypothesis is used, it may occur that for small n some of the β_{nk} lie outside of the domain of definition of $f(z)$; thus $r_n(z)$ need not be defined for sufficiently small n, but nevertheless is defined for n sufficiently large.

* The notation C_1 and C_2 is exceptional to this, but no confusion should arise; there is double notation for both C_1 and C_2.

those conditions the corresponding factors $z-\alpha_{nk}$ are simply to be omitted; and all of our formulas and conclusions remain correct without other modification.

We have hitherto (op. cit.) employed systematically condition (1) with the restriction $A_n=1$. Nevertheless the present more general form has appeared on occasion (for example, op. cit., pp. 206, 235, 261, 266, 274–275) and requires in the proof of Theorem 1 no change in method over the simpler form with $A_n=1$. Throughout the present paper we shall adopt (1) as standard.*

Even some elementary situations that are of interest are included in Theorem 1 but are not included if we require $A_n=1$. For instance we may take C_1 and C_2 as the circles $|z|=r_1$ and $|z|=r_2<r_1<r_0$, the numbers $\alpha_{n1}, \cdots, \alpha_{nn}$ as the nth roots of an arbitrary a_n whose modulus is not less than r_0^n, and the numbers $\beta_{n1}, \cdots, \beta_{n,n+1}$ as the $(n+1)$st roots of an arbitrary b_n whose modulus is not greater than r_2^{n+1}. Equation (1) is valid with $A_n=a_n$, independently of the behavior of a_n and b_n satisfying the conditions given; but equation (1) is not valid with $A_n=1$ unless the numbers $|a_n|^{1/n}$ approach a finite limit.

For the truth of (2) itself we assume $f(z)$ analytic throughout the interior of C_γ, but need not assume $f(z)$ to be analytic throughout the interior of no $C_{\gamma'}$ with $\gamma'>\gamma$. Indeed $f(z)$ may be analytic throughout the closed interior of C_{γ_1}. For our later purposes in the present paper, however, we find it desirable to make the complete assumption of Theorem 1. For appropriate examples illustrating convergence and degree of convergence when $f(z)$ is analytic throughout R, the reader may refer to the book already mentioned, page 239.

An interesting complement to Theorem 1 is

THEOREM 2. *Under the hypothesis of Theorem* 1 *we have for arbitrary* μ, $\gamma<\mu\leqq\gamma_1$,

$$\limsup_{n\to\infty} [\max\ |\ r_n(z)\ |,\ z \text{ on } C_\mu]^{1/n} \leqq \mu/\gamma. \tag{4}$$

* Contrary to the situation involving asymptotic conditions for poles, there would seem to be no advantage in the study of asymptotic conditions *for points* β_{nk} *of interpolation* in replacing the condition that

$$\lim_{n\to\infty} |(z-\beta_{n1})\cdots(\beta-\beta_{n,n+1})|^{1/n} \tag{a}$$

should exist uniformly by the condition that $\lim_{n\to\infty} |B_n(z-\beta_{n1})\cdots(z-\beta_{n,n+1})|^{1/n}$ should exist uniformly and not vanish identically; for it follows from Theorem 8 with $\alpha_{nk}=\infty$ that each of these conditions implies the other—this entire remark is made on the assumption that (a) is studied in the usual geometric situation, exterior to a curve within which the β_{nk} lie. On the other hand, such a relation as (34) is ordinarily considered interior to a curve to which the α_{nk} are exterior.

A formula for $r_n(z)$ is (op. cit., p. 186)

$$r_n(z) = \frac{1}{2\pi i}\int_\Gamma \left[1 - \frac{(z-\beta_{n1})\cdots(z-\beta_{n,n+1})}{(z-\alpha_{n1})\cdots(z-\alpha_{nn})} \cdot \frac{(t-\alpha_{n1})\cdots(t-\alpha_{nn})}{(t-\beta_{n1})\cdots(t-\beta_{n,n+1})}\right]\frac{f(t)dt}{t-z}, \qquad z \neq \alpha_{nk}, \tag{5}$$

where Γ is a Jordan curve on and within which $f(z)$ is analytic, and which contains all the points β_{nk} in its interior. Equation (5) is valid for all finite values of z other than the α_{nk}, even exterior to Γ, provided of course the integrand when not defined for such a value of z is replaced by its limit for that value of z.

Choose the numbers μ_1, μ_2, and μ_3 with $\mu_1 > \mu > \gamma > \mu_2 > \mu_3 > \gamma_2$, and choose Γ in (5) as the locus C_{μ_2}. For z on C_μ and t on C_{μ_2} we have by (1) when n is sufficiently large

$$\left| A_n \frac{(z-\beta_{n1})\cdots(z-\beta_{n,n+1})}{(z-\alpha_{n1})\cdots(z-\alpha_{nn})} \right| \leqq \mu_1^n,$$

$$\left| \frac{(t-\alpha_{n1})\cdots(t-\alpha_{nn})}{A_n(t-\beta_{n1})\cdots(t-\beta_{n,n+1})} \right| \leqq 1/\mu_3^n.$$

From (5) we read off at once

$$\limsup_{n\to\infty} [\max \mid r_n(z) \mid, z \text{ on } C_\mu]^{1/n} \leqq \mu_1/\mu_3,$$

and by allowing μ_1 to approach μ and μ_3 to approach γ we obtain (4).

2. **Degree of convergence.** We shall shortly obtain inequalities in opposite senses to (2) and (4), but in order to do this it is important to show how a certain degree of convergence on C_2 of rational functions $F_n(z)$ with poles in the points α_{nk} implies a corresponding degree of convergence on C_λ. The difficulty here lies in using the relation (1) directly, for the function

$$F_n(z) \div A_n \frac{(z-\beta_{n1})\cdots(z-\beta_{n,n+1})}{(z-\alpha_{n1})\cdots(z-\alpha_{nn})}$$

may have poles on C_2, and cannot be used for immediate comparison.* Nevertheless we shall prove

THEOREM 3. *Under the hypothesis of Theorem* 1 *let us suppose*

$$\limsup_{n\to\infty} [\max \mid F_n(z) \mid, z \text{ on } C_2]^{1/n} = q, \tag{6}$$

* Condition (44) is a consequence of (1) and applies here directly. Nevertheless condition (1) is more elementary and more natural; it seems desirable to prove Theorem 3 without using as intermediary Theorem 12, whose proof involves a different order of ideas.

where $F_n(z)$ is a rational function of degree n with its poles in the points α_{nk}. Then we have ($\gamma_2 < \lambda \leqq \gamma_1$)

$$\limsup_{n\to\infty} [\max \mid F_n(z) \mid, z \text{ on } C_\lambda]^{1/n} \leqq \lambda q/\gamma_2. \tag{7}$$

Let $q_1 > q$ be arbitrary, so that we have for n sufficiently large $\mid F_n(z) \mid \leqq q_1^n$, z on C_2. Let us denote by $w = \phi(z)$ a function which maps the exterior of C_2 onto $\mid w \mid > 1$ so that the points at infinity correspond to each other, and denote generically by J_N, $N > 1$, the locus $\mid \phi(z) \mid = N$ exterior to C_2. All the points α_{nk} lie exterior to a suitably chosen J_A, so for $Z < A$, $Z > 1$ we have (op. cit., p. 250, Lemma I)*

$$\mid F_n(z) \mid \leqq q_1^n \left[\frac{AZ - 1}{A - Z}\right]^n, \qquad z \text{ on } J_Z.$$

By the principle of maximum for analytic functions we have for z on and exterior to J_Z

$$\left| F_n(z) \div \left\{ A_n \frac{(z - \beta_{n1}) \cdots (z - \beta_{n,n+1})}{(z - \alpha_{n1}) \cdots (z - \alpha_{nn})} \right\} \right|$$

$$\leqq q_1^n \left[\frac{AZ - 1}{A \div Z}\right]^n \div \left[\min \left| A_n \frac{(z - \beta_{n1}) \cdots (z - \beta_{n,n+1})}{(z - \alpha_{n1}) \cdots (z - \alpha_{nn})} \right|, z \text{ on } J_Z \right],$$

$$\limsup_{n\to\infty} [\max \mid F_n(z) \mid, z \text{ on } C_\lambda]^{1/n} \leqq \lambda q_1 \frac{AZ - 1}{A - Z} \div [\min \Phi(z), z \text{ on } J_Z].$$

When we allow Z to approach unity, the curve J_Z approaches C_2, so by allowing q_1 to approach q we obtain (7).

Our most important preliminary result is now available:

THEOREM 4. *Under the hypothesis of Theorem* 1 *there exists no sequence of rational functions $R_n(z)$ of respective degrees n with poles in the points α_{nk} such that we have either of the relations*

$$\limsup_{n\to\infty} [\max \mid f(z) - R_n(z) \mid, z \text{ on } C_2]^{1/n} = \kappa/\gamma, \qquad \kappa < \gamma_2, \tag{8}$$

or ($\gamma_2 < \lambda < \gamma$)

$$\limsup_{n\to\infty} [\max \mid f(z) - R_n(z) \mid, z \text{ on } C_\lambda]^{1/n} = \kappa/\gamma, \qquad \kappa < \lambda. \tag{9}$$

Consequently, in Theorem 1 *the equality sign holds in both* (2) *and* (3).

* In the extension (§11) of Theorem 3 to a set C_2 composed of several Jordan curves, this inequality for $F_n(z)$ is established on a level curve J_Z corresponding to each of the Jordan curves; the remainder of the proof holds without change.

If we assume (8) or (9) to hold, we find by means of (3) or (2) for a specific λ, $\gamma_2 \leqq \lambda < \gamma$,

$$\limsup_{n\to\infty} [\max \mid r_n(z) - R_n(z) \mid, z \text{ on } C_\lambda]^{1/n} \leqq \lambda/\gamma,$$

whence from Theorem 3 for arbitrary μ, $\gamma < \mu \leqq \gamma_1$,

$$\limsup_{n\to\infty} [\max \mid r_n(z) - R_n(z) \mid, z \text{ on } C_\mu]^{1/n} \leqq \mu/\gamma;$$

a change of notation is necessary in this application of Theorem 3 if $\lambda > \gamma_2$. By Theorem 2 we now derive ($\gamma < \mu \leqq \gamma_1$)

$$\limsup_{n\to\infty} [\max \mid R_n(z) \mid, z \text{ on } C_\mu] \leqq \mu/\gamma. \tag{10}$$

The function $\log \Phi(z)$ is harmonic at every point of R, as the uniform limit of the sequence of harmonic functions

$$\frac{1}{n} \log \left| A_n \frac{(z-\beta_{n1}) \cdots (z-\beta_{n,n+1})}{(z-\alpha_{n1}) \cdots (z-\alpha_{nn})} \right|, \tag{11}$$

and is continuous in $\overline{R}$, equal to $\log \gamma_1$ and $\log \gamma_2$ on C_1 and C_1 respectively.* In any closed simply connected region interior to R, suitably chosen conjugates of the functions (11) converge uniformly to a suitably chosen conjugate of the function $\log \Phi(z)$. As z traces a curve C_λ in the counterclockwise sense, the conjugate of (11), which is the argument (that is, angle or amplitude) of the function

$$\left[\frac{(z-\beta_{n1}) \cdots (z-\beta_{n,n+1})}{(z-\alpha_{n1}) \cdots (z-\alpha_{nn})}\right]^{1/n},$$

increases by $2\pi(n+1)/n$; so as z traces C_λ the conjugate of $\log \Phi(z)$ increases by 2π. Consequently $\Phi(z)$ is not identically constant, and we have $\gamma_1 \neq \gamma_2$.

Theorem 4 is a consequence of the following theorem, a treatment of which the writer hopes to publish shortly in these Transactions. The theorem may be proved from the two-constant theorem, in a manner similar to that previously used by the present writer.† Theorem 5 is much more general than we need at the moment, and will be applied also later in the present paper.

THEOREM 5. *Let S be a region bounded by two disjoint Jordan curves K_0 and*

* We cannot have $\gamma_2 = 0$, for in that case the analytic function $\exp [\log \Phi(z) + i\Psi(z)]$, where $\Psi(z)$ is conjugate to $\log \Phi(z)$ in R, would approach the boundary value zero everywhere on C_2, and would be locally single-valued and analytic in R, hence (op. cit., §1.9) would vanish identically.

† Proceedings of the National Academy of Sciences, vol. 24 (1938), pp. 477–486; these Transactions, vol. 46 (1939), pp. 46–65.

K_{-1}, with K_{-1} interior to K_0. Let the function $u(x, y)$ be harmonic in S, continuous in the corresponding closed region $\overline{S}$, equal to zero and -1 on K_0 and K_{-1} respectively. Denote generically by K_σ the locus $u(x, y) = \sigma$, $0 > \sigma > -1$, and by S_σ the open set $\sigma > u(x, y) > -1$ in S bounded by K_σ and K_{-1}; denote by $\overline{S}_\sigma$ the closure of S_σ.

Let the function $f(z)$ be analytic throughout S_ρ but not analytic throughout any $S_{\rho'}$, $\rho' > \rho$, and let $f(z)$ be continuous in the two-dimensional sense on K_{-1} with respect to the domain $\overline{S}$. Let the function $f_n(z)$ be analytic in S, continuous in $\overline{S}$, with the relations

$$\limsup_{n\to\infty} [\max \mid f_n(z) \mid, z \text{ on } K_0]^{1/n} \leqq e^\alpha > 1, \tag{12}$$

$$\limsup_{n\to\infty} [\max \mid f(z) - f_n(z) \mid, z \text{ on } K_{-1}]^{1/n} \leqq e^\beta < 1. \tag{13}$$

Then we must have

$$\alpha + \alpha\rho - \beta\rho \geqq 0; \tag{14}$$

if the equality sign holds here we have

$$\limsup_{n\to\infty} [\max \mid f_n(z) \mid, z \text{ on } K_\mu]^{1/n} = e^{(\alpha-\beta)(\mu-\rho)}, \qquad 0 \geqq \mu \geqq \rho, \tag{15}$$

$$\limsup_{n\to\infty} [\max \mid f(z) - f_n(z) \mid, z \text{ on } K_\sigma]^{1/n} = e^{(\alpha-\beta)(\sigma-\rho)}, \qquad \rho > \sigma \geqq -1. \tag{16}$$

If Q is an arbitrary continuum in $\overline{S}$ not a single point, and if the equality sign holds in (14), *we have*

$$\limsup_{n\to\infty} [\max \mid f_{n+1}(z) - f_n(z) \mid, z \text{ on } Q]^{1/n} = e^{(\alpha-\beta)(U-\rho)}, \tag{17}$$

where $U = \max [u(x, y) \text{ on } Q]$. Consequently if the second member of (17) *is greater than unity, the first member is equal to*

$$\limsup_{n\to\infty} [\max \mid f_n(z) \mid, z \text{ on } Q]^{1/n}; \tag{18}$$

if the second member of (17) *is less than unity, the first member is equal to*

$$\limsup_{n\to\infty} [\max \mid f(z) - f_n(z) \mid, z \text{ on } Q]^{1/n}. \tag{19}$$

It is a consequence of (17) *and* (18) *that the sequence $f_n(z)$ converges throughout no region containing in its interior a point of K_μ, $0 > \mu \geqq \rho$.*

We apply Theorem 5 to the situation of Theorem 4 by identifying C_1 and C_λ ($\gamma_2 \leqq \lambda < \gamma$) with K_0 and K_{-1} respectively. Inequalities (10) for $\mu = \gamma_1$ and (8) or (9) are identified with (12) and (13), so we have

$$\alpha = \log \gamma_1 - \log \gamma, \qquad \beta = \log \kappa - \log \gamma.$$

We also set

$$u(x, y) = -\frac{\log \Phi(z) - \log \gamma_1}{\log \lambda - \log \gamma_1}, \qquad \rho = -\frac{\log \gamma - \log \gamma_1}{\log \lambda - \log \gamma_1}.$$

Direct substitution yields for (14) the inequality

$$\frac{\log \gamma - \log \gamma_1}{\log \lambda - \log \gamma_1} (\log \kappa - \log \lambda) \geqq 0,$$

whence $\kappa \geqq \lambda$, and Theorem 4 is established.

Let the equality sign hold in (14), under the conditions of Theorem 5. If the sequence $f_n(z)$ converges throughout some region containing in its interior a point of K_μ, $0 \geqq \mu \geqq \rho$, it follows, from Osgood's theorem to the effect that in a region of convergence subregions of uniform convergence are everywhere dense, that the sequence $f_n(z)$ converges uniformly in some closed region Q containing in its interior a point of some $K_{\mu'}$, $0 > \mu' > \rho$, contrary to the equality of (18) with the second member of (17).

The sequence $f_n(z)$ can converge like a convergent geometric series on no continuum in $\overline{S}$ exterior to K_ρ and consisting of more than one point.

3. **Maximal convergence.** Theorem 4 is our chief justification for the

DEFINITION. *Under the hypothesis of Theorem* 1, *any sequence of rational functions* $R_n(z)$ *of respective degrees* n *with poles in the points* $\alpha_{n1}, \alpha_{n2}, \cdots, \alpha_{nn}$ *is said to converge maximally to* $f(z)$ *on the set* C *constituting the closed interior of* C_2 *provided we have*

$$\limsup_{n \to \infty} [\max \mid f(z) - R_n(z) \mid, z \text{ on } C]^{1/n} = \gamma_2/\gamma. \tag{20}$$

We mention explicitly that maximal convergence is not defined if $f(z)$ is analytic throughout the interior of C_1.

As an immediate consequence of the definition we have from Theorems 1 and 4

THEOREM 6. *Under the hypothesis of Theorem* 1, *the sequence* $r_n(z)$ *converges maximally to* $f(z)$ *on* C.

Of course the condition

$$\limsup_{n \to \infty} [\max \mid f(z) - R_n(z) \mid, z \text{ on } C_\lambda]^{1/n} \leqq \lambda/\gamma,$$

holding for all λ greater than but sufficiently near γ_2, is sufficient for maximal convergence, for we have

$$[\max \mid f(z) - R_n(z) \mid, z \text{ on } C_2] \leqq [\max \mid f(z) - R_n(z) \mid, z \text{ on } C_\lambda],$$

and after making the corresponding substitution we may allow λ to approach γ_2.

As an alternative in the definition of maximal convergence we may replace (20) by

$$\limsup_{n\to\infty} [\max \mid R_{n+1}(z) - R_n(z) \mid, z \text{ on } C]^{1/n} \leqq \gamma_2/\gamma, \tag{21}$$

provided that the sequence $R_n(z)$ is assumed to converge to $f(z)$ on C; under this assumption we may also replace (20) by (21) with the sign $\leqq$ replaced by the equality sign. This remark is an immediate consequence of the easily proved relation

$$\limsup_{n\to\infty} [\max \mid R_{n+1}(z) - R_n(z) \mid, z \text{ on } C]^{1/n} = \limsup_{n\to\infty} [\max \mid f(z) - R_n(z) \mid, z \text{ on } C]^{1/n},$$

provided either of these expressions is less than unity.

THEOREM 7. *Under the hypothesis of Theorem* 1, *let* $R_n(z)$ *be a sequence of rational functions of respective degrees* n *whose poles lie in the points* α_{nk}, *and which converges maximally to* $f(z)$ *on* C. *Then we have* $(\gamma_2 \leqq \lambda < \gamma,\ \gamma \leqq \mu \leqq \gamma_1)$

$$\limsup_{n\to\infty} [\max \mid f(z) - R_n(z) \mid, z \text{ on } C_\lambda]^{1/n} = \lambda/\gamma, \tag{22}$$

$$\limsup_{n\to\infty} [\max \mid R_n(z) \mid, z \text{ on } C_\mu]^{1/n} = \mu/\gamma. \tag{23}$$

If Q *is an arbitrary continuum in* $\overline{R}$ *not a single point, we have*

$$\limsup_{n\to\infty} [\max \mid R_{n+1}(z) - R_n(z) \mid, z \text{ on } Q]^{1/n} = [\max \Phi(z), z \text{ on } Q]/\gamma. \tag{24}$$

Consequently if the second member of (24) *is less than unity we have*

$$\limsup_{n\to\infty} [\max \mid f(z) - R_n(z) \mid, z \text{ on } Q]^{1/n} = [\max \Phi(z), z \text{ on } Q]/\gamma; \tag{25}$$

if the second member of (24) *is greater than or equal to unity we have*

$$\limsup_{n\to\infty} [\max \mid R_n(z) \mid, z \text{ on } Q]^{1/n} = [\max \Phi(z), z \text{ on } Q]/\gamma. \tag{26}$$

It is a consequence of (26) *that the sequence* $R_n(z)$ *converges throughout no region containing in its interior a point of* C_μ, $\gamma_1 > \mu \geqq \gamma$.

From the assumed maximal convergence of $R_n(z)$ and from the maximal convergence of $r_n(z)$ (Theorem 6) we have

$$\limsup_{n\to\infty} [\max \mid r_n(z) - R_n(z) \mid, z \text{ on } C_2]^{1/n} \leqq \gamma_2/\gamma,$$

whence from Theorem 3

$$\limsup_{n\to\infty} [\max \mid r_n(z) - R_n(z) \mid, z \text{ on } C_1]^{1/n} \leqq \gamma_1/\gamma.$$

By application of Theorem 2 we may now write

$$\limsup_{n\to\infty} [\max \mid R_n(z) \mid, z \text{ on } C_1]^{1/n} \leqq \gamma_1/\gamma. \tag{27}$$

Inequalities (27) and (20) with C replaced by C_2 place us in a position to apply Theorem 5 again. We identify C_1 and C_2 with K_0 and K_{-1} respectively, and we set from (27) and (20)

$$\alpha = \log \gamma_1 - \log \gamma, \qquad \beta = \log \gamma_2 - \log \gamma.$$

Moreover we have

$$u(x, y) = -\frac{\log \Phi(z) - \log \gamma_1}{\log \gamma_2 - \log \gamma_1}, \qquad \rho = -\frac{\log \gamma - \log \gamma_1}{\log \gamma_2 - \log \gamma_1}.$$

Direct computation shows that the equality sign is valid in (14), and hence the conclusion of Theorem 7 follows.

Theorem 7 is a generalization and a sharpening of the corresponding result (op. cit., §§4.7 and 4.8) for the case of approximation by polynomials; relations (24), (25), (26) are new even in the latter case. For approximation by polynomials we have $\alpha_{nk} = \infty$; according to the usual convention (op. cit., §§8.1, 8.2, 8.5) the corresponding factors $z - \alpha_{nk}$ in (1) are simply to be omitted. With $\alpha_{nk} = \infty$ there exist (op. cit., chap. 7) various sets of points β_{nk} for which the relation (1) obtains; for instance we may take $\Phi(z) = \Delta \mid \phi(z) \mid$, where $w = \Delta \cdot \phi(z)$ maps the exterior of C_2 onto $\mid w \mid > \Delta$ with $\phi'(\infty) = 1$; the first such set β_{nk} was exhibited by Fejér, uniformly distributed on C_2 with respect to a suitably chosen parameter.

In connection with Theorem 7 it is worth remarking that maximal convergence of the sequence $R_n(z)$ to $f(z)$ on the closed interior of C_2 implies the maximal convergence of the sequence $R_n(z)$ to $f(z)$ on the closed interior of every C_λ, $\gamma_2 < \lambda < \gamma$.

4. **Completion of a partial sequence of α_{nk}.** Theorem 1 is valid if the α_{nk} and β_{nk}, and hence also the functions $r_n(z)$, are defined not for every n but merely for an infinite sequence of indices n. But in the proof of Theorem 7 we have employed both $R_n(z)$ and $R_{n+1}(z)$, and thus have made essential use of the fact that the $R_n(z)$ are defined for every n. It is sufficient for our purposes thus far, even in the study of maximal convergence, as the reader may

notice, if the α_{nk} and $R_n(z)$ are defined not for every n but for an infinite sequence of indices n_j, with $n_{j+1}>n_j$ and $n_{j+1}-n_j$ bounded for all j. But if the α_{nk} and $R_n(z)$ are defined only for a sequence of indices n_j, and if $n_{j+1}-n_j$ is not bounded, our fundamental conclusions (22) and (23) may fail even for the specific functions $R_n(z)\equiv r_n(z)$ of Theorem 1, as we now proceed to show by examples. These examples, chosen from Taylor's series, are closely related to gap theorems and to overconvergence in the sense of Ostrowski.

Choose $\alpha_{nk}=\infty$, $\beta_{nk}=0$, so that we have $\Phi(z)\equiv|z|$. Choose C_2 as the circle $|z|=1/2$, C_γ as the circle $|z|=1$, and C_1 as the circle $|z|=2$. Choose the integer n_1 so that $2^{n_1-1}>3$, the integer $n_2>n_1$ so that $2^{(n_2-1)/n_1}>3$, and in general the integer $n_{k+1}>n_k$ so that $2^{(n_{k+1}-1)/n_k}>3$. It is sufficient to choose $n_0=0$, $n_1=3$, $n_{k+1}=2n_k+1$. If we set

$$f(z) = 1 + z^{n_1} + z^{n_2} + z^{n_3} + \cdots, \tag{28}$$

and denote by $r_n(z)$ the sum of the first $n+1$ terms of the corresponding series with all of the powers of z present by indication, we have by Theorem 7

$$\limsup_{n\to\infty} [\max\,|f(z)-r_n(z)|,\ \text{for}\ |z|=1/2]^{1/n}=1/2. \tag{29}$$

But we have for $|z|=1/2$

$$|f(z)-r_{n_k}(z)| \leqq \frac{1}{2^{n_{k+1}}}+\frac{1}{2^{n_{k+2}}}+\cdots<\frac{1}{2^{n_{k+1}-1}},$$

whence

$$\limsup_{k\to\infty} [\max\,|f(z)-r_{n_k}(z)|,\ \text{for}\ |z|=1/2]^{1/n_k} \leqq \limsup_{k\to\infty}\frac{1}{2^{(n_{k+1}-1)/n_k}} \leqq \frac{1}{3},$$

in contrast to (29).

With the same choice of α_{nk}, β_{nk}, C_2, C_1, and C_γ, let us now choose the integer n_1 so that $2^{1/(n_1-1)}<3/2$, the integer $n_2>n_1$ so that $2^{(n_1+1)/(n_2-1)}<3/2$, and in general the integer n_{k+1} so that $2^{(n_k+1)/(n_{k+1}-1)}<3/2$; it is sufficient to choose $n_0=0$, $n_{k+1}=2n_k+3$. Again we define $f(z)$ by equation (28), and we denote by $r_n(z)$ the sum of the first $n+1$ terms of the corresponding series with all the powers of z indicated. From Theorem 7 we have

$$\limsup_{n\to\infty} [\max\,|r_n(z)|,\ \text{for}\ |z|=2]^{1/n}=2. \tag{30}$$

But from (28) we may write for $|z|=2$

$$|r_{n_k-1}(z)| \leqq 1+2^{n_1}+2^{n_2}+\cdots+2^{n_k-1} \leqq 2^{n_k-1+1},$$

whence

$$\limsup_{k\to\infty} [\max \mid r_{n_k-1}(z) \mid, \text{ for } \mid z \mid = 2]^{1/(n_k-1)}$$

$$\leqq \limsup_{k\to\infty} 2^{(n_{k-1}+1)/(n_k-1)} \leqq 3/2,$$

in contrast to (30).

The examples just formulated, in connection with such inequalities as (2), (3), and (4), suggest the following problem. Let C_1, C_2, R, $\Phi(z)$ be given so that the conditions of Theorem 1 are satisfied, including equation (1) but where the α_{nk} and β_{nk} are defined not for all n, only for an infinite sequence of indices; naturally, in equation (1) only those indices are admitted. *To define now new numbers α_{nk} and β_{nk} where necessary so that equation* (1) *shall hold for all n.* We proceed now to the discussion and solution of this problem. In the solution we need the following preliminary result:

THEOREM 8. *Let the relation for some infinite sequence of indices n*

$$\lim_{n\to\infty} \left| A_n \frac{(z-\beta_{n1}) \cdots (z-\beta_{n,n+1})}{(z-\alpha_{n1}) \cdots (z-\alpha_{nn})} \right|^{1/n} = e^{U(x,y)} \tag{31}$$

be valid uniformly on every closed set interior to an annular region R bounded by Jordan curves C_1 and C_2, with C_2 interior to C_1, where each β_{nk} lies on or interior to C_2 and each α_{nk} lies on or exterior to C_1. Then the function $U(x, y)$ is harmonic in R and not identically constant in R; indeed we have*

$$\int_\Gamma \frac{\partial U}{\partial \nu}\, ds = 2\pi, \tag{32}$$

where ν indicates the exterior normal, where Γ is an arbitrary analytic Jordan curve separating C_1 and C_2, and where integration is in the counterclockwise sense. At every finite point exterior to C_2 we have

$$\lim_{n\to\infty} \mid (z-\beta_{n1}) \cdots (z-\beta_{n,n+1}) \mid^{1/n}$$

$$= \exp\left[\frac{1}{2\pi}\int_{\Gamma_2}\left(U\frac{\partial \log r}{\partial \nu} - \log r \frac{\partial U}{\partial \nu}\right) ds\right], \qquad z = x + iy, \tag{33}$$

uniformly on any closed finite set exterior to C_2, where Γ_2 is an analytic Jordan curve in R containing C_2 but not (x, y) in its interior; at every point (x, y) interior to C_1 we have

---

* The left-hand member of (31) if existent uniformly on any closed set in R vanishes at every point of R or at no point of R. For the corresponding analytic functions are locally single-valued, different from zero, and form a normal family in any simply connected subregion of R. By Hurwitz's theorem any limit function of the family vanishes at every point or at no point of R.

$$\lim_{n\to\infty} \left| (z-\alpha_{n1}) \cdots (z-\alpha_{nn})/A_n \right|^{1/n}$$

$$= \exp\left[\frac{-1}{2\pi}\int_{\Gamma_1}\left(U\frac{\partial \log r}{\partial \nu} - \log r \frac{\partial U}{\partial \nu}\right)ds\right], \qquad z = x + iy, \tag{34}$$

uniformly on any closed set interior to C_1, where Γ_1 is an analytic Jordan curve in R containing (x, y) and C_2 but not C_1 in its interior. In (33) *and* (34) *the limits are to be taken over the given sequence of indices for which* (31) *holds. The integrals over Γ_1 and Γ_2 are to be taken in the counterclockwise and clockwise senses respectively, and ν denotes exterior and interior normal respectively.*

Theorem 8 is only a slight modification of a previous result (Theorem 18, op. cit., p. 266), and the proof is therefore left to the reader. In the direction of a converse we have

THEOREM 9. *Let R denote an annular region bounded by Jordan curves C_1 and C_2 with C_2 interior to C_1. Let the function $U(x, y)$ be harmonic interior to R, continuous in the corresponding closed region, taking the values g_1 and $g_2 < g_1$ on C_1 and C_2 respectively, and satisfying equation* (32), *where Γ is an arbitrary analytic Jordan curve separating C_1 and C_2. Then for every n there exist points α'_{nk} on C_1 and points β'_{nk} on C_2 and constants A_n such that we have*

$$\lim_{n\to\infty} \left| A_n \frac{(z-\beta'_{n1}) \cdots (z-\beta'_{n,n+1})}{(z-\alpha'_{n1}) \cdots (z-\alpha'_{nn})} \right|^{1/n} = e^{U(x,y)} \tag{35}$$

uniformly on any closed set interior to R.

If we replace A_n in equation (35) by $q^n A_n$, where q is positive, the equation persists with $U(x, y)$ replaced by $U(x, y) + \log q$. In particular, then, it is no loss of generality to assume $g_1 = 0$ in the proof of Theorem 9; we make this assumption.

To exhibit the points α'_{nk} and β'_{nk} desired it is now sufficient to choose the α'_{nk} and β'_{nk} uniformly distributed on C_1 and C_2 respectively with respect to the parameter σ, where

$$d\sigma = \frac{\partial U}{\partial \nu} ds \text{ on } C_1, \qquad d\sigma = -\frac{\partial U}{\partial \nu} ds \text{ on } C_2,$$

for we have the equation for (x, y) in R

$$U(x, y) = \frac{1}{2\pi}\int_{C_2} \log r \, d\sigma - \frac{1}{2\pi}\int_{C_1} \log r \, d\sigma. \tag{36}$$

These equations for $d\sigma$ and $U(x, y)$ still have a meaning in an extended sense even if the Jordan curves C_1 and C_2 are not analytic, and so also does the con-

cept of uniform distribution of points on C_1 and C_2; compare op. cit., §§7.6 and 9.12. Further details of the proof of Theorem 9 are so similar to a proof given elsewhere (Theorem 9, op. cit., pp. 210–211) that they are omitted here. Equation (35) is valid with $A_n=1$ and with $U(x, y)$ replaced by $U(x, y)-g_1$, hence is valid in the original form with $A_n=e^{n\sigma_1}$.

We are now in a position to solve the proposed problem:

THEOREM 10. *Under the hypothesis of Theorem* 8, *where* $U(x, y)$ *takes constant values on* C_1 *and* C_2, *new numbers* α_{nk} *and* β_{nk} *can be defined for those values of* n *for which the original* α_{nk} *and* β_{nk} *are not employed in* (31), *in such a way that* (31) *holds uniformly on any closed set interior to* R *for the entire sequence* $n=1, 2, 3, \cdots$.

The original sequences α_{nk} and β_{nk} used in (31) yield by Theorem 8 a function $U(x, y)$ which assumes constant values g_1 and g_2 on C_1 and C_2, and which satisfies (32), where Γ is an arbitrary analytic Jordan curve separating C_1 and C_2, and where ν indicates exterior normal. It then follows that we have $g_2<g_1$. By virtue of Theorem 9 there exist for every n points α'_{nk} and β'_{nk} which satisfy (35), uniformly on any closed set interior to R. For the values of n that do not appear in the sequence in the relation (31) of our hypothesis we now define $\alpha_{nk}=\alpha'_{nk}$, $\beta_{nk}=\beta'_{nk}$ for those values of n. Then the points α_{nk} and β_{nk} are now defined for every n, and equation (31) is valid uniformly on any closed set interior to R for the complete sequence $n=1, 2, 3, \cdots$.

Theorem 10 applies directly in the situation of Theorem 1 but where (1) is assumed merely for a suitable sequence of indices n. But it is to be noticed that in Theorem 1 equation (1) is assumed to hold uniformly on C_1, whereas in Theorem 10 equation (1) holds uniformly merely on any closed set interior to R.

Proof of Theorem 10 by means of Theorems 8 and 9 is essentially the execution of a program previously outlined (op. cit., p. 268 ff.).

5. **Determination of the β_{nk} when C and the α_{nk} are given.** Our definition of maximal convergence involves the points α_{nk}, the point set C, and the function $\Phi(z)$, but does not involve the points β_{nk} directly; of course the β_{nk} are intimately related to the function $\Phi(z)$. This raises the question of the determination of $\Phi(z)$ and the β_{nk} when C and the α_{nk} are given, a question that we proceed to discuss.

By Theorem 8, condition (34) is a consequence of (31), so (34) or some similar relation is the natural hypothesis for us to use on the points α_{nk}. Such a condition as (34) is fulfilled whenever the points α_{nk} are uniformly distributed on a Jordan curve with respect to a continuous parameter.

THEOREM 11. *Let the point set* C *be the closed interior of a Jordan curve* C_2.

Let C_1' be a Jordan curve containing in its interior the set C but none of the points α_{nk}, and let us suppose

$$\lim_{n\to\infty} \left| (z-\alpha_{n1})\cdots(z-\alpha_{nn})/A_n \right|^{1/n} = e^{U(x,y)} \tag{37}$$

interior to C_1', uniformly on any closed set interior to C_1'. Then there exists a region R bounded by C_2 and by a Jordan curve C_1 containing C_2 in its interior, and there exists a function $V(x, y)$ harmonic in R, continuous in the corresponding closed region, constant on C_1 and on C_2; for suitably chosen points β_{nk} on C_2 we have interior to R

$$\lim_{n\to\infty} \left| A_n \frac{(z-\beta_{n1})\cdots(z-\beta_{n,n+1})}{(z-\alpha_{n1})\cdots(z-\alpha_{nn})} \right|^{1/n} = e^{V(x,y)}, \tag{38}$$

uniformly on any closed set interior to R.

The function $U(x, y)$ can be written as the limit of the sequence of harmonic functions

$$U_n(x, y) \equiv (1/n) \log \left| (z-\alpha_{n1})\cdots(z-\alpha_{nn})/A_n \right|$$

uniformly on any closed set interior to C_1', so $U(x, y)$ is harmonic interior to C_1'.

Let $G(x, y)$ denote Green's function for the complement K of C with pole at infinity; thus $G(x, y) - \frac{1}{2} \log (x^2+y^2)$ is harmonic at infinity, and $G(x, y)$ vanishes on C_2. Let $W(x, y)$ denote the unique function which is harmonic in K (even at infinity), continuous in the corresponding closed region, and equal to $U(x, y)$ on C_2. Introduce the notation

$$V'(x, y) = W(x, y) + G(x, y) - U(x, y), \tag{39}$$

so that $V'(x, y)$ is harmonic in the annular region R' bounded by C_1' and C_2, and is continuous with the value zero on C_2. If Γ denotes an arbitrary analytic Jordan curve separating C_1' and C_2, and if ν denotes exterior normal, we have

$$\int_\Gamma \frac{\partial V'}{\partial \nu}\, ds = \int_\Gamma \frac{\partial W}{\partial \nu}\, ds + \int_\Gamma \frac{\partial G}{\partial \nu}\, ds - \int_\Gamma \frac{\partial U}{\partial \nu}\, ds = 2\pi; \tag{40}$$

for the first and third integrals of the second member vanish because $W(x, y)$ and $U(x, y)$ are harmonic respectively in the closed exterior and closed interior of Γ.

We shall now show* that $V'(x, y)$ *is positive at every point of* R'. Consider

* The entire proof of Theorem 11 can be interpreted as the carrying out in detail of a method previously indicated (op. cit., p. 269). But the inequality $V'>0$ is an indispensable condition for the validity of that method, and the present proof of that inequality is the first ever given of wide generality.

the function (compare (39))

$$V_n'(x, y) = W(x, y) + G(x, y) - U_n(x, y),$$

which approaches $V'(x, y)$ uniformly on any closed set in $R'+C_2$. The function $V_n'(x, y)$ is defined and harmonic at every point exterior to C_2, except the points α_{nk} and except at infinity. When (x, y) approaches a finite point $z=\alpha_{nk}$, the functions $-U_n(x, y)$ and $V_n'(x, y)$ become positively infinite. When (x, y) becomes infinite, the functions $W(x, y)$, $G(x, y)-\frac{1}{2}\log(x^2+y^2)$, and $\frac{1}{2}\log(x^2+y^2)-U_n(x, y)$ approach finite limits, except that $\frac{1}{2}\log(x^2+y^2)-U_n(x, y)$ becomes positively infinite if one or more of the points α_{nk} are infinite. Consequently the minimum of the function $V_n'(x, y)$ considered in the entire closed region exterior to C_2 exists and occurs for (x, y) on the curve C_2 itself:

$$[V_n'(x, y), \text{ for } (x, y) \text{ exterior to } C_2] \geqq [\min V_n'(x, y), \text{ on } C_2].$$

By the uniformity of convergence of $V_n'(x, y)$ to $V'(x, y)$ on C_2 as n becomes infinite, this right-hand member approaches the minimum of $V'(x, y)$ on C_2; thus

$$[V'(x, y), \text{ for } (x, y) \text{ in } R'] \geqq [\min V'(x, y), \text{ on } C_2].$$

Consequently for (x, y) in R' we have proved $V'(x, y)\geqq 0$. The function $V'(x, y)$ is zero on C_2, can approach no negative value as (x, y) in R' approaches C_1', and is harmonic in the region bounded by C_1' and C_2. But $V'(x, y)$ is not identically zero in R', by equation (40). It follows from the well known properties of the minima of harmonic functions that the inequality $V'(x, y) > 0$ persists throughout R'.

We shall now exhibit the desired points β_{nk}. Let us suppose for the moment that C_2 is an analytic Jordan curve. For (x, y) exterior to C_2 we have (op. cit., p. 266, Lemma IV)

$$W(x, y) + G(x, y) = \frac{1}{2\pi}\int_{C_2}\left[(W+G)\frac{\partial \log r}{\partial \nu} - \log r \frac{\partial (W+G)}{\partial \nu}\right] ds + q, \tag{41}$$

where ν indicates the interior normal for C_2, and where q is suitably chosen. For (x, y) exterior to C_2 we also have (op. cit., p. 265, Lemma III)

$$0 = \frac{1}{2\pi}\int_{C_2}\left(U\frac{\partial \log r}{\partial \nu} - \log r \frac{\partial U}{\partial \nu}\right) ds;$$

this equation is first established with U replaced by U_n, and then use is made

of the uniformity of convergence of U_n and $\partial U_n/\partial \nu$ on C_2. It follows by subtraction from (41) for (x, y) exterior to C_2 that

$$W(x, y) + G(x, y) = \frac{1}{2\pi}\int_{C_2}\left(V' \frac{\partial \log r}{\partial \nu} - \log r \frac{\partial V'}{\partial \nu}\right) ds + q$$
$$= \frac{-1}{2\pi}\int_{C_2} \log r \frac{\partial V'}{\partial \nu} ds + q. \tag{42}$$

Thanks to the inequality $V'(x, y) > 0$ in R', we have on C_2 the inequality $d\sigma = -(\partial V'/\partial \nu)ds \geqq 0$. If now the points β_{nk} are chosen uniformly distributed on C_2 with respect to the parameter σ, we may write from (42) and from $-\int_{C_2}(\partial V'/\partial \nu)ds = 2\pi$ (a consequence of (40)) the relation

$$\lim_{n\to\infty} |(z - \beta_{n1}) \cdots (z - \beta_{n,n+1})|^{1/n} = e^{W(z,y)+G(z,y)-q} \tag{43}$$

for z exterior to C_2, uniformly on any closed set exterior to C_2 (compare op. cit., §8.7). It follows now from (37) and (43) by the choice $V(x, y) = V'(x, y) - q$ that equation (38) is valid uniformly on any closed set interior to R'.

In (41) and later equations we have for the sake of convenience assumed C_2 to be an analytic Jordan curve. It is sufficient if C_2 is an arbitrary Jordan curve, provided the integrals are interpreted in an extended sense (compare op. cit., §7.6). Even if C_2 is a still more general set, the integrals may be taken over analytic Jordan curves near but exterior to C_2 on which $V(x, y)$ is constant (which does not alter the validity of (42) or of the other equations), and the points β_{nk} may be chosen on these curves by the method of op. cit., §4.4; but here we must relax the requirement that the β_{nk} shall lie on or interior to C_2, and may replace it by the requirement that the β_{nk} shall have no limit point exterior to C_2.

A region R satisfying the requirements of Theorem 11 is easily defined. Let g_1 denote the least upper bound of $V'(x, y)$ in R', and let g be arbitrary, $0 < g < g_1$. Then the locus C_1: $V'(x, y) = g$ in R' is an analytic Jordan curve; the annular region bounded by C_1 and C_2 fulfills all the conditions of Theorem 11. Our fundamental results on approximation apply only to loci $V'(x, y) = \text{const.}$, so it is no great disadvantage to cut R off along such a curve.

It is not uninteresting to note that if in Theorem 11 we assume equation (37) to hold merely for an infinite sequence of indices n, then (as in Theorem 10) new points α_{nk} can be provided, and also the points β_{nk}, so that both the α_{nk} and β_{nk} shall be defined for every n with (38) valid. Indeed, the definition of $U(x, y)$ and the proof of (43) do not assume the α_{nk} defined for every n; Theorem 11 can be used to determine the β_{nk} for the values of n for which

the α_{nk} are defined, and then Theorem 10 applies. Of course equation (43) is substantially the same as equation (33) of Theorem 8.

6. **Invariant formulation of results.** The general problem of greatest degree of approximation to a given analytic function $f(z)$ by rational functions with prescribed poles α_{nk}, culminating in Theorem 7, is invariant under linear transformation of the complex variable. But our fundamental condition (1) is not invariant, and has no obvious invariant properties. We proceed to discuss conditions analogous to (1) but having properties of invariance.

THEOREM 12. *Under the conditions of Theorem* 1 *(it is sufficient if* (1) *holds uniformly merely on any closed set in R) let the function* $w=\phi(z)$ *map the exterior of* C_2 *onto* $|w|>1$ *so that the points at infinity correspond to each other. Then we have*

$$\lim_{n\to\infty}\left|\frac{[\bar{\phi}(\alpha_{n1})\phi(z)-1]\cdots[\bar{\phi}(\alpha_{nn})\phi(z)-1]}{[\phi(z)-\phi(\alpha_{n1})]\cdots[\phi(z)-\phi(\alpha_{nn})]}\right|^{1/n}=\Phi_1(z), \tag{44}$$

uniformly on any closed set in R, where $\Phi_1(z)$ *is a suitable constant multiple of the function* $\Phi(z)$ *of* (1).

From condition (1) it follows that we have (34) fulfilled; it follows from (34) (method of op. cit., §9.12) that

$$\lim_{n\to\infty}\left|[\phi(z)-\phi(\alpha_{n1})]\cdots[\phi(z)-\phi(\alpha_{nn})]/A_n\right|^{1/n}$$

exists uniformly on any closed set interior to R. By another method previously employed (op. cit., §9.4) it follows that (44) is valid uniformly on any closed set in R, where $\Phi_1(z)$ is suitably chosen. It is to be noted that the function whose absolute value appears in (44) can be considered defined even on C_2, and to be continuous there, whence $\Phi_1(z)$ also is continuous on C_2: $\Phi_1(z)\equiv 1$ on C_2. Then the limit in (44) can be considered uniform on any closed set in $\bar{R}$ containing no point of C_1. The function whose absolute value appears in (44) is greater than unity at every point exterior to C_2, so we have $\Phi_1(z)\geqq 1$ in R.

From (1) and (44) we may write

$$\lim_{n\to\infty}\left|A_n\frac{(z-\beta_{n1})\cdots(z-\beta_{n,n+1})}{(z-\alpha_{n1})\cdots(z-\alpha_{nn})}\cdot\frac{[\phi(z)-\phi(\alpha_{n1})]\cdots[\phi(z)-\phi(\alpha_{nn})]}{[\bar{\phi}(\alpha_{n1})\phi(z)-1]\cdots[\bar{\phi}(\alpha_{nn})\phi(z)-1]}\right|^{1/n}=\frac{\Phi(z)}{\Phi_1(z)}, \tag{45}$$

uniformly on any closed set in R. Denote by $\Psi_n(z)(z-\beta_{n1})$ the function whose modulus occurs in (45). Then $\Psi_n(z)$ is analytic and different from zero at

every point of the extended plane exterior to C_2, when the function is suitably defined in the points α_{nk}.* Moreover $\lim_{n\to\infty}|z-\beta_{n1}|^{1/n}=1$ uniformly on any closed bounded set exterior to C_2. Then (45), being valid on an arbitrary C_γ, $\gamma_1>\gamma>\gamma_2$ when suitably interpreted, is valid at every finite point in the closed exterior of C_γ; the logarithm of the right-hand member of (45) can be considered defined and harmonic at every point of the extended plane exterior to C_2, is continuous in the corresponding closed region, and has the constant value $\log \gamma_2$ on C_2. Hence throughout R we have $\Phi(z)/\Phi_1(z)\equiv\gamma_2$, and Theorem 12 is established.

Under the conditions of Theorem 12 it may occur that C_1 is a curve $|\phi(z)|=\text{const.}$, say c; under such circumstances the function

$$\log \Phi(z) - \frac{\log \gamma_1 - \log \gamma_2}{\log c} \log |\phi(z)| - \log \gamma_2$$

is harmonic in R, continuous in $\overline{R}$, zero on C_1 and C_2, hence is zero in $\overline{R}$:

$$\log \Phi(z) = \frac{\log \gamma_1 - \log \gamma_2}{\log c} \log |\phi(z)| + \log \gamma_2.$$

But by (32) we have

$$\int_\Gamma \frac{\partial \log \Phi(z)}{\partial \nu}\, ds = 2\pi,$$

where Γ is an analytic Jordan curve separating C_1 and C_2; also (op. cit., p. 71)

$$\int_\Gamma \frac{\partial \log |\phi(z)|}{\partial \nu}\, ds = 2\pi,$$

whence we deduce $(\log \gamma_1-\log \gamma_2)/\log c=1$, and for z in $\overline{R}$,

$$\Phi(z) \equiv \gamma_2 |\phi(z)|.$$

This equation is satisfied, it may be added, provided the points α_{nk} and β_{nk} are distributed on C_1 and C_2 respectively uniformly with respect to the harmonic function conjugate to $\log |\phi(z)|$; see op. cit., §4.3.

The converse of Theorem 12 can also be established:

THEOREM 13. *Let R be the region between and bounded by two Jordan curves C_1 and C_2, with C_2 interior to C_1. Let $w=\phi(z)$ map the exterior of C_2 onto the*

* When a point α_{nk} lies at infinity, the function $[\phi(z)-\phi(\alpha_{nk})]/[\overline{\phi}(\alpha_{nk})\phi(z)-1]$ in (44) and (45) is to be replaced by its limit as $\alpha_{nk}\to\infty$, namely $1/\phi(z)$. In accordance with the convention already made (§1) concerning factors $z-\alpha_{nk}$, there is here no exception in the behavior of $\Psi_n(z)$ even if points α_{nk} are infinite.

region $|w|>1$ *so that the points at infinity correspond to each other. Let the points* α_{nk} *lie exterior to* C_1, *and suppose* (44) *valid uniformly on any closed set in* $\overline{R}$ *not containing a point of* C_1. *Then for suitably chosen numbers* A_n *and suitably chosen points* β_{nk} *on* C_2, *equation* (1) *is valid uniformly on any closed set in* R, *where* $\Phi(z)$ *is a constant multiple of* $\Phi_1(z)$.

From equation (44) follows (op. cit., p. 274, Corollary 2) an equation of type (37), valid uniformly on any closed set interior to C_1. By Theorem 11 we now derive a condition of form (1), and inspection of the proof of Theorem 11 shows that equation (1) is valid uniformly on any closed set in R. It follows now from Theorem 12 that $\Phi(z)$ is a constant multiple of $\Phi_1(z)$, so Theorem 13 is established.

In Theorem 13 the Jordan curve C_1 naturally need not be a locus obtained by setting $\Phi(z)$ equal to a constant.

We remark (op. cit., p. 274) that condition (44), valid uniformly on any closed set in R, implies the existence of

$$\lim_{n\to\infty}\left|\frac{(z-\alpha_{n1})\cdots(z-\alpha_{nn})}{\phi(\alpha_{n1})\cdots\phi(\alpha_{nn})}\right|^{1/n}$$

uniformly on any closed set in R; if a particular α_{nk} is infinite, the corresponding quotient $(z-\alpha_{nk})/\phi(\alpha_{nk})$ is to be replaced by the limit of that quotient as α_{nk} becomes infinite, namely $-1/\phi'(\infty)$. By virtue of (33), a consequence of Theorems 13 and 8, it now follows that equation (1) is valid uniformly on any closed set in R for a suitably chosen $\Phi(z)$ provided we have

$$A_n=\phi(\alpha_{n1})\cdots\phi(\alpha_{nn}). \tag{46}$$

The proof just given, that (44) implies the existence of

$$\lim_{n\to\infty}\left|\phi(\alpha_{n1})\cdots\phi(\alpha_{nn})\,\frac{(z-\beta_{n1})\cdots(z-\beta_{n,n+1})}{(z-\alpha_{n1})\cdots(z-\alpha_{nn})}\right|^{1/n},$$

shows the uniform existence of this limit merely on any closed set in R; the limit is different from zero there. But if (1) itself is satisfied under the conditions of Theorem 1, it then follows that for the original A_n

$$\lim_{n\to\infty}\left|\frac{\phi(\alpha_{n1})\cdots\phi(\alpha_{nn})}{A_n}\right|^{1/n}$$

exists and is different from zero. Consequently (1) with the substitution (46) made is valid uniformly on any closed set in $\overline{R}$ containing no point of C_2. It is not essentially more general to consider (1) in its original form than to consider (1) with (46) satisfied.

Let us now consider the function $\Psi_n(z)$ defined in connection with (45), where A_n is given by (46). We see by inspection that $\Psi_n(\infty)=1$, from which it follows that $\Phi(z)/\Phi_1(z)$ is identically unity. Thus we have proved

COROLLARY 1. *Under the conditions of Theorem* 1, *we have*

$$\lim_{n\to\infty}\left|\phi(\alpha_{n1})\cdots\phi(\alpha_{nn})\,\frac{(z-\beta_{n1})\cdots(z-\beta_{n,n+1})}{(z-\alpha_{n1})\cdots(z-\alpha_{nn})}\right|^{1/n}=\Phi_1(z), \tag{47}$$

uniformly on any closed set in $\overline{R}$ *containing no point of* C_2, *where* $\Phi_1(z)$ *is given by* (44).

If z_0 is an arbitrary point interior to C_1, the hypothesis of Theorem 8 implies (34) for $z=z_0$, so for the given sequence of indices

$$\lim_{n\to\infty}\left|(z_0-\alpha_{n1})\cdots(z_0-\alpha_{nn})/A_n\right|^{1/n}$$

exists. This limit is different from zero, by the form of (34) itself. It follows now from (31) that for the given sequence of indices

$$\lim_{n\to\infty}\left|\frac{(z_0-\alpha_{n1})\cdots(z_0-\alpha_{nn})(z-\beta_{n1})\cdots(z-\beta_{n,n+1})}{(z-\alpha_{n1})\cdots(z-\alpha_{nn})}\right|^{1/n}$$

exists uniformly on any closed set in R. Consequently whenever equation (1) is satisfied, that equation is also satisfied for some function $\Phi(z)$ with

$$A_n=(z_0-\alpha_{n1})\cdots(z_0-\alpha_{nn}),$$

where z_0 is an arbitrary point interior to C_1.

As a matter of record we formulate

COROLLARY 2. *Under the hypothesis of Theorem* 13, *let* $R_n(z)$ *be a sequence of rational functions of respective degrees* n, *whose poles lie in the points* α_{nk}. *Denote generically by* C_γ *the curve* $\Phi_1(z)=\gamma$ *in* R, *and let the function* $f(z)$ *be analytic throughout the interior of* C_γ *but not throughout the interior of any* $C_{\gamma'}$, $\gamma'>\gamma$.

A sufficient condition for the maximal convergence of the sequence $R_n(z)$ *to* $f(z)$ *on* C (*closed interior of* C_2) *is*

$$\limsup_{n\to\infty}\left[\max\left|f(z)-R_n(z)\right|, z \text{ on } C_2\right]^{1/n}\leqq 1/\gamma; \tag{48}$$

if the inequality sign holds here, so also does the equality sign.

Condition (44) is invariant under linear transformation of the complex variable, and in some respects is therefore more advantageous than condition (1). It is obviously immaterial whether the loci C_γ and the right-hand

members of (22)–(26) are determined from $\Phi(z)$ or from $\Phi_1(z)$; in the future we shall use these functions interchangeably.

The method of proof of Theorem 12, namely the use of such an equation as (45), yields

COROLLARY 3. *Under the conditions of Theorem* 1, *suppose also that points* β'_{nk} *exist on or within the curve* C_γ, $\gamma_1 > \gamma \geqq \gamma_2$, *so that the relation*

$$\lim_{n\to\infty} \left| A_n' \frac{(z-\beta'_{n1})\cdots(z-\beta'_{n,n+1})}{(z-\alpha_{n1})\cdots(z-\alpha_{nn})} \right|^{1/n} = \Phi_2(z) \not\equiv \text{const.}$$

holds uniformly on any closed set in $\overline{R}'$ *containing no point of* C_γ, *where* R' *is the region between and bounded by* C_1 *and* C_γ, *and where* $\Phi_2(z)$ *is constant on* C_γ. *Then the quotient* $\Phi_2(z)/\Phi(z)$ *is identically constant in* R'.

We remark too that this corollary can be established by the use of Theorem 7 itself.

7. **Maximal convergence of extremal sequences.** We are now in a position expeditiously to treat extremal sequences.

THEOREM 14. *Under the conditions of Theorem* 1, *let* C *denote the closed interior of* C_2 *and let* $R_n(z)$ *be the (or a) rational function of degree* n *with poles in the points* α_{nk}, *of best approximation to* $f(z)$ *on* C *in the sense of Tchebycheff, or in the sense of least pth powers* $(p>0)$ *over* C_2 *(assumed rectifiable), or in the sense of least pth powers* $(p>0)$ *as measured by a surface integral over* C, *or in the sense of least pth powers over* γ: $|w|=1$ *when* K *is mapped onto the exterior of* γ *so that the points at infinity correspond to each other, or in the sense of least pth powers over the circumference* $|w|=1$ *or over the closed region* $|w|\leqq 1$ *when* C *is mapped onto* $|w|\leqq 1$, *in every case with a positive continuous weight function. Then the sequence* $R_n(z)$ *converges maximally to* $f(z)$ *on* C.

Inequality (48) is established for the present sequence $R_n(z)$ precisely as in previous cases (op. cit., p. 264, Theorem 17; p. 254, Theorem 12). The details are left to the reader.

The restriction that the weight function be positive and continuous can be considerably lightened in the various cases; compare op. cit., §5.7.

8. **Necessary conditions on the β_{nk} for uniform convergence.** We have hitherto presented such conditions as (1) and (47) as sufficient conditions on the β_{nk} for maximal convergence. We now consider their necessity as conditions for uniform and maximal convergence.

THEOREM 15. *Let* C_1 *and* C_2 *be Jordan curves, with* C_2 *interior to* C_1, *and let* R *denote the annular region between them. Let the points* α_{nk} *lie exterior to* C_1 *and*

satisfy the condition

$$\lim_{n\to\infty}\left|\frac{[\overline{\phi(\alpha_{n1})}\phi(z)-1]\cdots[\overline{\phi(\alpha_{nn})}\phi(z)-1]}{[\phi(z)-\phi(\alpha_{n1})]\cdots[\phi(z)-\phi(\alpha_{nn})]}\right|^{1/n}=\Phi(z) \tag{49}$$

uniformly in $\overline{R}$; *such a condition as* (49) *is satisfied whenever the condition* (37) *is fulfilled uniformly on and within* C_1. *Suppose the points* β_{nk} *are given on or within* C_2, *and suppose the corresponding sequence of interpolating rational functions* $r_n(z)$ *of respective degrees* n *converges uniformly to* $f(z)$ *on and within* C_2 *whenever* $f(z)$ *is analytic on and within* C_2. *Then we have uniformly on any closed set in* R *the equation*

$$\lim_{n\to\infty}\left|\phi(\alpha_{n1})\cdots\phi(\alpha_{nn})\frac{(z-\beta_{n1})\cdots(z-\beta_{n,n+1})}{(z-\alpha_{n1})\cdots(z-\alpha_{nn})}\right|^{1/n}=\Phi(z). \tag{50}$$

Consequently the sequence $r_n(z)$ *converges maximally to* $f(z)$ *on the closed interior of* C_2 *whenever* $f(z)$ *is analytic in that closed region but not analytic throughout* R.

This theorem is due to Kalmár for the case $\alpha_{nk}=\infty$; the present method is due to the present writer (op. cit., §7.3), an extension of the method in the case $\alpha_{nk}=\infty$.

We introduce the notation

$$\Psi_n(z)=\phi(\alpha_{n1})\cdots\phi(\alpha_{nn})\frac{(z-\beta_{n1})\cdots(z-\beta_{n,n+1})}{\phi(z)(z-\alpha_{n1})\cdots(z-\alpha_{nn})};$$

it is obvious that $\lim_{n\to\infty}|\phi(z)|^{1/n}=1$ uniformly on any closed limited set exterior to C_2, so (50) is equivalent to

$$\lim_{n\to\infty}|\Psi_n(z)|^{1/n}=\Phi(z) \tag{51}$$

uniformly on any closed set in R. The relation $\Phi(z)\geqq 1$ in R follows at once from (49).

The sequence

$$\left|\frac{(z-\alpha_{n1})\cdots(z-\alpha_{nn})}{\phi(\alpha_{n1})\cdots\phi(\alpha_{nn})}\right|^{1/n}$$

converges uniformly on any closed set in R to a nonvanishing function (op. cit., p. 274). Hence the sequence $\log|\Psi_n(z)|^{1/n}$ is uniformly bounded on any closed set in R, and forms a normal family of harmonic functions in R;* these

* We are using here easily proved properties of sequences of harmonic and analytic functions, uniformly bounded and hence equicontinuous on any closed set interior to R.

We also use the fact that a uniformly convergent sequence of nonvanishing analytic functions converges to a nonvanishing analytic function unless the limit function vanishes identically.

functions are obviously single-valued and harmonic in $\bar{R}$. Denote by $\log|\Psi(z)|$ an arbitrary limit function of the family. We shall prove the

LEMMA. *Either* $|\Psi(z)|$ *approaches the constant unity as* z *in* R *approaches* C_2 *or for some points of* R *the function* $|\Psi(z)|$ *has a value greater than unity and for other points of* R *this function* $|\Psi(z)|$ *has a value less than unity.*

A subsequence of the functions $|\Psi_n(z)|^{1/n}$ approaches $|\Psi(z)|$ uniformly on any closed subset of R, so the corresponding sequence

$$\left|\Psi_n(z)\frac{[\phi(z)-\phi(\alpha_{n1})]\cdots[\phi(z)-\phi(\alpha_{nn})]}{[\overline{\phi(\alpha_{n1})}\phi(z)-1]\cdots[\overline{\phi(\alpha_{nn})}\phi(z)-1]}\right|^{1/n} \tag{52}$$

approaches $|\Psi(z)|/\Phi(z)$ uniformly on any closed subset of R. The function whose modulus appears in (52) is analytic and different from zero exterior to C_2, even in the points α_{nk} when properly defined there, and has the modulus $1/|\phi'(\infty)|$ at infinity. The sequence (52) therefore converges uniformly on any closed set of the extended plane exterior to C_2, to the modulus of a function analytic exterior to C_2 and which has unit modulus at infinity. Thus $\log|\Psi(z)/\Phi(z)|$ has a meaning and is harmonic in the extended plane exterior to C_2, and has the value zero at infinity, even though $\Phi(z)$ and $\Psi(z)$ are not properly defined exterior to C_1.

Of course each of the quantities in (49) whose limit is $\Phi(z)$ is greater than unity in R; the function $\Phi(z)$ is not constant in R, so we have $\Phi(z)>1$ in R (compare op. cit., p. 229); but $\Phi(z)$ is continuous and equal to unity on C_2. The lemma now follows from the well known properties of the maxima and minima of harmonic functions.

We introduce the notation $\omega_n(z)=\phi(z)\Psi_n(z)$. In the expansion of the function $1/(t-z)$, where t is exterior to C_2, by interpolating functions $r_n(z)$, we have (op. cit., §8.1)

$$f(z)-r_n(z)=\frac{\omega_n(z)}{\omega_n(t)(t-z)}\cdot \tag{53}$$

If we introduce the notation

$$M_n=\max\,[|\omega_n(z)|,\ z \text{ on } C_2], \tag{54}$$

our hypothesis implies

$$\lim_{n\to\infty} M_n/\omega_n(t)=0. \tag{55}$$

If (51) is not satisfied, the normality of the family $|\Psi_n(z)|^{1/n}$ implies that some limit function $|\Psi(z)|$ of this family is different from $\Phi(z)$. Then for

some point t, which may be chosen in R, we have by the lemma for suitably chosen indices n_k

$$\lim_{k\to\infty} |\Psi_{n_k}(t)|^{1/n_k} = \Psi(t) < 1.$$

We may also write

$$\lim_{k\to\infty} |\omega_{n_k}(t)|^{1/n_k} = \Psi(t) < 1,$$

and for suitably large index

$$|\omega_{n_k}(t)| < 1. \tag{56}$$

Choose this point t as the value in (53). From the definition of $\omega_n(z)$ and from the properties of $\phi(z)$ on C_2 we have

$$M_n = \max [|\omega_n(z)|, z \text{ on } C_2] = \max [|\Psi_n(z)|, z \text{ on } C_2].$$

From the properties at infinity of the function whose modulus appears in (52) we have

$$\max [|\Psi_n(z)|, z \text{ on } C_2] \geqq 1/|\phi'(\infty)|, \qquad M_n \geqq 1/|\phi'(\infty)|. \tag{57}$$

Inequalities (56) and (57) are in contradiction to (55), so equation (51) is established. The remainder of Theorem 15 is a consequence of Theorem 6.

A result closely related to Theorem 15, an extension of a result due to Fekete (op. cit., p. 163), is

THEOREM 16. *Let C_1, C_2, R, and the α_{nk} satisfy the conditions of Theorem* 15, *including* (49). *Let the points β_{nk} lie on or within C_2, and let M_n be defined by* (54). *Then*

$$\lim_{n\to\infty} M_n^{1/n} = 1 \tag{58}$$

is a necessary and sufficient condition that the sequence of rational functions $r_n(z)$ of respective degrees n with poles in the points α_{nk} and interpolating to $f(z)$ in the points β_{nk}, should converge uniformly to $f(z)$ on and within C_2 whenever $f(z)$ is analytic on and within C_2. Thus (58) *is also a necessary and sufficient condition for the maximal convergence of the sequence $r_n(z)$ to $f(z)$ whenever $f(z)$ is analytic on and within C_2 but is not analytic throughout R.*

If condition (58) is satisfied, the functions

$$\left[\Psi_n(z)\frac{[\phi(z)-\phi(\alpha_{n1})]\cdots[\phi(z)-\phi(\alpha_{nn})]}{[\overline{\phi(\alpha_{n1})}\phi(z)-1]\cdots[\overline{\phi(\alpha_{nn})}\phi(z)-1]}\right]^{1/n} \tag{59}$$

are locally single-valued and analytic in the extended plane exterior to C_2 and

form a normal family in that region in an extended sense (op. cit., p. 162). No limit function of the family can have a modulus in that region greater than unity, by (58). But each function of the family has a value at infinity of modulus $|\phi'(\infty)|^{-1/n}$, and each limit function has a value at infinity of modulus unity. Consequently every limit function of the family is of modulus unity throughout the extended plane exterior to C_2, and the sequence (52) approaches unity uniformly on any closed set exterior to C_2. Then by (49) equation (50) is valid uniformly on any closed set in R, which implies the uniform convergence of $r_n(z)$ to $f(z)$ whenever $f(z)$ is analytic on and within C_2, and the maximal convergence under these conditions if $f(z)$ is not analytic throughout R.

Conversely, let us now assume that the sequence $r_n(z)$ converges uniformly to $f(z)$ on and within C_2 whenever $f(z)$ is analytic on and within C_2. If (58) does not hold, we have from the reasoning used on the functions (59) for a suitable infinite sequence n_k and for a suitable $\Delta > 1$

$$M_{n_k} \geqq \Delta^{n_k} > 1. \tag{60}$$

By Theorem 15 equation (50) is satisfied. Choose t in R with $1 < \Phi(t) < \Delta$. By (51) we have for sufficiently large n

$$|\omega_n(t)| \leqq \Delta^n. \tag{61}$$

For the function $f(z) = 1/(t-z)$ inequalities (60) and (61) contradict our hypothesis (55). This contradiction proves (58), and implies the maximal convergence of $r_n(z)$ to $f(z)$ whenever $f(z)$ is analytic on and within C_2 but not analytic throughout R. Theorem 16 is established.

It should be mentioned that results closely related to Theorems 15 and 16 have been previously established by Shen (op. cit., p. 258); but the hypothesis on the α_{nk} is there geometric, not asymptotic, and the necessary and sufficient conditions derived, such as (50) and (58), are not for uniform convergence, but for an inequality corresponding to (2).

Theorems 15 and 16 are of interest chiefly in connection with maximal convergence. Nevertheless under the conditions of both those theorems (compare the remarks made concerning Theorem 1), whenever $f(z)$ is analytic throughout the interior of $\Phi(z) = T$ in R, we have

$$\limsup_{n\to\infty} [\max |f(z) - r_n(z)|, z \text{ on } C_2]^{1/n} \leqq 1/T.$$

In connection with Theorems 15 and 16 we shall prove the

COROLLARY. *Let C_1, C_2, R, and the α_{nk} satisfy the conditions of Theorem* 15, *including* (49). *Let the points β_{nk} lie on C_2. Then a necessary and sufficient con-*

dition that the sequence of rational functions $r_n(z)$ of respective degrees n with poles in the points α_{nk} and interpolating to $f(z)$ in the points β_{nk} should converge uniformly to $f(z)$ on and within C_2 whenever $f(z)$ is analytic on and within C_2, is that the points β_{nk} be uniformly distributed on C_2 with respect to the function conjugate to $\log \Phi(z)$ *in R, continuous in $\overline{R}$.*

The sufficiency of this condition (due to Fejér if $\alpha_{nk} = \infty$) is essentially contained in the proof of Theorem 11, in the light of the proof of Theorem 13. To prove the necessity of the condition, which is due to Kalmár if $\alpha_{nk} = \infty$, we note that (50) is a consequence of Theorem 15 and that (33) then follows; in (33) we may take Γ_2 as identical with C_2 if the integral is interpreted in an extended sense; by the fact that $U(x, y)$ is constant on C_2 we may write (33) in the form

$$\lim_{n\to\infty} \left| (z - \beta_{n1}) \cdots (z - \beta_{n,n+1}) \right|^{1/n} = \exp\left[\frac{-1}{2\pi} \int_{C_2} \log r \, \frac{\partial U}{\partial \nu} \, ds \right],$$

uniformly on any closed bounded set exterior to C_2. The conclusion follows at once by methods previously set forth (op. cit., §7.6).

On the topic of uniform distribution of points it is appropriate to establish

THEOREM 17. *Let C_1, C_2, R, and the α_{nk} satisfy the conditions of Theorem* 8, *including* (31), *and let $U(x, y)$ be constant on C_1. Let the points α_{nk} lie on C_1. Then the α_{nk} are uniformly distributed on C_1 with respect to the function conjugate to $U(x, y)$ in R, continuous in $\overline{R}$.*

Let z_0 be an arbitrary fixed point interior to C_1. In (34) we take the integral (in an extended sense) over C_1; by the fact that $U(x, y)$ is constant on C_1 we have

$$\lim_{n\to\infty} \left| \frac{(z - \alpha_{n1}) \cdots (z - \alpha_{nn})}{(z_0 - \alpha_{n1}) \cdots (z_0 - \alpha_{nn})} \right|^{1/n} = \exp\left[\frac{1}{2\pi} \int_{C_1} \log \frac{r}{r^0} \cdot \frac{\partial U}{\partial \nu} \, ds \right],$$

uniformly for z on any closed set interior to C_1, where α is a variable on C_1 and $r = |z - \alpha|$, $r^0 = |z_0 - \alpha|$. If we set

$$du = \frac{1}{2\pi} \frac{\partial U}{\partial \nu} \, ds,$$

we may then write

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} \log \frac{r_{nk}}{r_{nk}^0} = \int_{C_1} \log \frac{r}{r^0} \cdot du,$$

$$r_{nk} = |z - \alpha_{nk}|, \qquad r_{nk}^0 = |z_0 - \alpha_{nk}|.$$

By virtue of (32), this integral over C_1 can be considered an integral over the interval $0 \leqq u \leqq 1$. An arbitrary function of α continuous on C_1 can be uniformly approximated on C_1 as closely as desired by a suitably chosen constant plus a linear combination of functions

$$\log [\,|\, z_k - \alpha \,|\, /\,|\, z_0 - \alpha \,|\,],$$

where the z_k are suitably chosen points interior to C_1; this follows by the transformation $z' = 1/(z_0 - \alpha)$ from a former lemma (op. cit., p. 169, Lemma II).* An arbitrary function $\chi(\alpha)$ constant on C_1 has the trivial property

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} \chi(\alpha_{nk}) = \int_{C_1} \chi(\alpha) du,$$

so it follows from the possibility of approximation that every function continuous on C_1 has the corresponding property. It follows (compare op. cit., §§7.5 and 7.6) that the set α_{nk} is uniformly distributed on C_1 with respect to the parameter u.

9. **Maximal convergence of sequences interpolating in the Fekete-Shen points.** Certain sets of points, first introduced by Fekete for the case of polynomials, were used by him as points of interpolation to define polynomials converging favorably to a given analytic function on a given point set. The analogous points for the case of rational functions with preassigned poles α_{nk} were introduced and used by Shen (op. cit., §9.7) when the α_{nk} are subject to geometric conditions; Shen's results on degree of convergence, even for the case $\alpha_{nk} = \infty$, are more precise than those of Fekete. These same points were used subsequently by the present writer (op. cit., §9.10) when the α_{nk} satisfy asymptotic conditions. We are now in a position to prove

THEOREM 18. *Let C_1, C_2, R, and the α_{nk} satisfy the conditions of Theorem* 15, *including* (49). *Let the β_{nk} be the Fekete-Shen points for C_2. Then conditions* (50) *and* (58) *are fulfilled. Consequently whenever $f(z)$ is analytic on and within C_2 but not analytic throughout R, the sequence of rational functions $r_n(z)$ of respective degree n with poles α_{nk} defined by interpolation to $f(z)$ in the points β_{nk} converges maximally to $f(z)$ on and within C_2.*

If $f(z)$ is analytic throughout the interior of the curve C_T: $\Phi(z) = T$ in R, we have (op. cit., p. 263, Corollary 2)

$$\limsup_{n\to\infty} [\max \,|\, f(z) - r_n(z) \,|\,, z \text{ on } C_2]^{1/n} \leqq 1/T.$$

* It is essential to admit an additive constant, for otherwise no nonvanishing constant could itself be approximated; a uniformly convergent sequence of approximating functions converges to a uniform limit in the closed exterior of C_1, and each function $\log |(z_k - \alpha)/(z_0 - \alpha)|$ is harmonic and vanishes at infinity: $\alpha = \infty$.

The maximal convergence of the sequence $r_n(z)$ follows from Corollary 2 to Theorem 13, if $f(z)$ is not analytic throughout the interior of any $C_{T'}$, $T'>T$. The remainder of Theorem 18 follows from Theorems 15 and 16.

It is of interest to note that the Fekete-Shen points β_{nk} lie on C_2 itself, and hence by the Corollary to Theorem 16 these points are uniformly distributed on C_2 with respect to the function conjugate to $\log \Phi(z)$ in R, continuous in $\overline{R}$.

It is of some interest to note that the entire theory of maximal convergence of sequences of rational functions may be developed by taking not (1) but condition (49) as fundamental, using the Fekete-Shen points (as in op. cit., §9.10) to obtain a sequence of rational functions $r_n(z)$ satisfying the above inequality; the analogue of (4) is readily proved, so that Theorem 5 applies. The remainder of the theory can then be built up. It seems simpler and more natural to the writer to make (1) fundamental rather than (49).

10. **A general extremal problem on approximation by functions with preassigned poles.** The problem of approximation by rational functions with preassigned poles can be considered a special case of approximation by more general functions with preassigned poles.

THEOREM 19. *Let C_1, C_2, R, and the points α_{nk} and β_{nk} satisfy the conditions of Theorem* 1, *including condition* (1). *Let R_γ denote generically the annular region bounded by C_2 and C_γ. Suppose the function $F(z)$ to be analytic in R_γ, $\gamma_1>\gamma>\gamma_2$, but not analytic throughout any $R_{\gamma'}$, $\gamma'>\gamma$; suppose $F(z)$ continuous in the two-dimensional sense on C_2.*

Denote by K the region exterior to C_2. Then there exist functions $F_n(z)$ analytic in K except for possible poles in the points α_{nk}, continuous in the two-dimensional sense on C_2, such that we have

$$\limsup_{n\to\infty} [\max \mid F(z) - F_n(z) \mid, z \text{ on } C_2]^{1/n} = \gamma_2/\gamma; \tag{62}$$

but there exists no such sequence $F_n(z)$ for which the left-hand member of (62) *is less than γ_2/γ.*

Any sequence $F_n(z)$ for which (62) *holds possesses also the following properties*:

$$\limsup_{n\to\infty} [\max \mid F(z) - F_n(z) \mid, z \text{ on } C_\lambda]^{1/n} = \lambda/\gamma, \qquad \gamma_2 \leqq \lambda < \gamma, \tag{63}$$

$$\limsup_{n\to\infty} [\max \mid F_n(z) \mid, z \text{ on } C_\mu]^{1/n} = \mu/\gamma, \qquad \gamma \leqq \mu \leqq \gamma_1. \tag{64}$$

If Q is an arbitrary continuum in $\overline{R}$ not a single point, we have

$$\limsup_{n\to\infty} [\max \mid F_{n+1}(z) - F_n(z) \mid, z \text{ on } Q]^{1/n} = [\max \Phi(z), z \text{ on } Q]/\gamma. \tag{65}$$

The function $F(z)$ can be written as the sum of its two components:

$$(66_1) \qquad F(z) \equiv f_1(z) + f_2(z), \qquad z \text{ in } R_\gamma,$$

$$(66_2) \qquad f_1(z) \equiv \frac{1}{2\pi i}\int_{\Gamma_1} \frac{F(t)dt}{t-z}, \qquad z \text{ interior to } C_\gamma,$$

$$(66_3) \qquad f_2(z) \equiv \frac{1}{2\pi i}\int_{\Gamma_2} \frac{F(t)dt}{t-z}, \qquad z \text{ exterior to } C_2,$$

where Γ_1 is a rectifiable Jordan curve in R_γ containing z and C_2 in its interior, and where Γ_2 is a rectifiable Jordan curve in R_γ containing C_2 in its interior but having z exterior to it; the curves Γ_1 and Γ_2 depend then on z, but the functions $f_1(z)$ and $f_2(z)$ are independent of the particular curves chosen. Equation (66_1) is valid for z in R_γ. Nevertheless the integrals in (66_2) and (66_3) define $f_1(z)$ and $f_2(z)$ as functions analytic respectively throughout the interior of C_γ and throughout the exterior of C_2 (even at infinity, by a limiting process). Equation (66_1) can be used to define $f_2(z)$ on C_2; with this additional definition, the function $f_2(z)$ is continuous on C_2. The function $f_1(z)$ is analytic throughout the interior of no $C_{\gamma'}$, $\gamma' > \gamma$.

We are now in a position to identify the function $f(z)$ of Theorem 1 with the present $f_1(z)$. If we write

$$f(z) \equiv f_1(z), \qquad r_n(z) \equiv F_n(z) - f_2(z),$$

equation (62) is a consequence of (indeed identical with) equation (3) (see Theorem 4).

The method of proof of Theorems 3 and 4 applies also in the present case, and shows that there exists no sequence $F_n(z)$ for which the left-hand member of (62) is less than γ_2/γ.

Let us suppose now $F_n(z)$ to be an arbitrary sequence satisfying (62). With the aid of the comparison sequence $r_n(z)+f_2(z)$ just considered, which satisfies (62) in place of $F_n(z)$, we find as in the proof of (10)

$$\limsup_{n\to\infty} [\max \, |F_n(z)|, z \text{ on } C_\mu]^{1/n} \leqq \mu/\gamma, \qquad \gamma < \mu \leqq \gamma_1.$$

Theorem 5 now applies, and yields the remaining parts of Theorem 19.

A slight modification of Theorem 19 is of interest. Let us replace the requirement that $F(z)$ be continuous in the two-dimensional sense on C_2 by the requirement that $F(z)$ be bounded in the neighborhood of C_2, with the identical modification in the requirements on $F_n(z)$. The expression

$$\max \, |F(z) - F_n(z)|, \quad z \text{ on } C_2,$$

which occurs in (62) is then to be interpreted as

$$\limsup_{\lambda \to \gamma_2} [\max \, | F(z) - F_n(z) |, z \text{ on } C_\lambda], \tag{67}$$

which is necessarily finite. With this understanding, also similarly for (63), (64), and (65), Theorem 5 is valid for the corresponding sequences, and Theorem 19 is valid in the generalized form.

With this new convention, there exists for each n at least one admissible function $F_n(z)$ for which (67) *is least.* By an admissible function $F_n(z)$ we understand here a function analytic in K except for possible poles in the points α_{nk}, and bounded in the neighborhood of C_2. Let N_0 denote the greatest lower bound of all numbers (67) for the class of admissible functions. Let $F_n^{(1)}(z)$, $F_n^{(2)}(z)$, $\cdots$ be admissible functions for which the corresponding numbers $N^{(1)}$, $N^{(2)}$, $\cdots$ approach N_0:

$$\lim_{k \to \infty} N^{(k)} = N_0. \tag{68}$$

We have

$$\begin{aligned} &[\max \, | F_n^{(1)}(z) - F_n^{(k)}(z) |, z \text{ on } C_\lambda] \\ &\leqq [\max \, | F(z) - F_n^{(1)}(z) |, z \text{ on } C_\lambda] + [\max \, | F(z) - F_n^{(k)}(z) |, z \text{ on } C_\lambda], \\ &\limsup_{\lambda \to \gamma_2} [\max \, | F_n^{(1)}(z) - F_n^{(k)}(z) |, z \text{ on } C_\lambda] \leqq N^{(1)} + N^{(k)}. \end{aligned} \tag{69}$$

The function $F_n^{(1)}(z) - F_n^{(k)}(z)$ is analytic in K except for poles in the points α_{nk}. The function

$$\frac{[F_n^{(1)}(z) - F_n^{(k)}(z)][\phi(z) - \phi(\alpha_{n1})] \cdots [\phi(z) - \phi(\alpha_{nn})]}{[\overline{\phi(\alpha_{n1})}\phi(z) - 1] \cdots [\overline{\phi(\alpha_{nn})}\phi(z) - 1]}, \tag{70}$$

when suitably defined in the points α_{nk}, is analytic throughout K and by (69) is uniformly bounded (independently of k) in K. From any infinite sequence of functions (70) can be extracted a subsequence which converges throughout K, uniformly on any closed set in K, to a bounded function of form (70) with $F_n^{(k)}(z)$ replaced by an admissible function $F_n^{(0)}(z)$. The corresponding subsequence of the $F_n^{(k)}(z)$ then converges to $F_n^{(0)}(z)$ throughout K except in the points α_{nk}, uniformly on any closed set in K containing no point α_{nk}. Let us suppose (this supposition involves merely a change of notation) that the original sequence $F_n^{(k)}(z)$ so converges.

Let μ be chosen arbitrarily but fixed, $\gamma > \mu > \gamma_2$, and let us suppose $\gamma_2 < \lambda < \mu$, where λ will be allowed to approach γ_2. We now apply the two-constant theorem (or Hadamard three-circle theorem), which is valid under

the present generalized conditions on the functions involved:

$$[\max\,|F(z)-F_n^{(k)}(z)|,\ z \text{ on } C_\lambda] \leqq [\max\,|F(z)-F_n^{(k)}(z)|,\ z \text{ on } C_2]^{(\mu-\lambda)/(\mu-\gamma_2)} \cdot [\max\,|F(z)-F_n^{(k)}(z)|,\ z \text{ on } C_\mu]^{(\lambda-\gamma_2)/(\mu-\gamma_2)}. \tag{71}$$

The second square bracket in the second member of (71) has a bound N independent of k; the first square bracket in the second member of (71) approaches N_0, by (68). Thus we have by letting k become infinite

$$[\max\,|F(z)-F_n^{(0)}(z)|,\ z \text{ on } C_\lambda] = \lim_{k\to\infty}[\max\,|F(z)-F_n^{(k)}(z)|,\ z \text{ on } C_\lambda] \leqq N_0^{(\mu-\lambda)/(\mu-\gamma_2)} \cdot N^{(\lambda-\gamma_2)/(\mu-\gamma_2)}; \tag{72}$$

the first equality sign is a consequence of the uniform convergence of $F_n^{(k)}(z)$ to $F_n^{(0)}(z)$ on C_λ. If we allow λ to approach γ_2 in (72), we have

$$[\max\,|F(z)-F_n^{(0)}(z)|,\ z \text{ on } C_2] \leqq N_0;$$

the inequality sign is impossible here by the definition of N_0, so we have established the existence of at least one extremal function $F_n(z)$. The question of the uniqueness of this extremal function is still open.

Naturally the extremal functions $F_n(z)$ whose existence has just been established satisfy condition (62); and for them the conclusion of Theorem 19 is valid.

We remark too that the requirements of continuity of $F(z)$ and $F_n(z)$ on C_2 need not be replaced by the requirement of boundedness in the neighborhood of C_2 but may simply be omitted, without essentially altering our conclusions; the square bracket in the left-hand member of (62) is merely understood as (67); a suitable comparison sequence $F_n(z)$ can be obtained as before from Theorem 1; at least one extreme function $F_n(z)$ exists.

Whether or not this last generalization is allowed, we shall prove

THEOREM 20. *Let the conditions of Theorem* 19 *be fulfilled, including equation* (62), *and let the respective components of the function* $F_n(z)$ *be* $f_{n1}(z)$ *and* $f_{n2}(z)$, *defined by equations analogous to* (66). *Then the function* $f_{n1}(z)$ *is a rational function of degree* n *with poles in the points* α_{nk}, *and the sequence* $f_{n1}(z)$ *converges maximally to* $f_1(z)$ *on the closed interior of* C_2.

In the region R_γ we have the equation

$$F_n(z) = f_{n1}(z) + f_{n2}(z),$$

where $f_{n1}(z)$ is analytic throughout the interior of C_γ, and $f_{n2}(z)$ is analytic

throughout the exterior of C_2, even at infinity. But $F_n(z)$ is analytic throughout K except in the points α_{nk}, so $f_{n1}(z)$ can be extended analytically from R_γ, and can be considered defined and analytic throughout K except for poles in the points α_{nk}. Consequently $f_{n1}(z)$ is a rational function of degree n whose poles lie in the points α_{nk}.

In the defining equations for $f_1(z)$ and $f_{n1}(z)$ we may choose Γ_1 as the curve C_λ, $\gamma_2<\lambda<\gamma$, whence

$$f_1(z) - f_{n1}(z) = \frac{1}{2\pi i}\int_{C_\lambda} \frac{F(t) - F_n(t)}{t - z}\, dt, \qquad z \text{ on } C_2.$$

Then equation (63) yields

$$\limsup_{n\to\infty} [\max \mid f_1(z) - f_{n1}(z) \mid, z \text{ on } C_2]^{1/n} \leqq \lambda/\gamma.$$

In this inequality we may allow λ to approach γ_2:

$$\limsup_{n\to\infty} [\max \mid f_1(z) - f_{n1}(z) \mid, z \text{ on } C_2]^{1/n} \leqq \gamma_2/\gamma. \tag{73}$$

The function $f_2(z)$ is analytic throughout the region K, and $F(z)$ is analytic throughout the interior of C_γ but not throughout the interior of any $C_{\gamma'}$, $\gamma'>\gamma$. Then $f_1(z)$ also is analytic throughout the interior of C_γ but not throughout the interior of any $C_{\gamma'}$. Inequality (73) thus establishes the conclusion of Theorem 20.

Results analogous to but less explicit than Theorems 19 and 20 lie at hand if we replace the original requirement on $F(z)$ by the requirement that $F(z)$ be analytic throughout R. For instance there exist functions $F_n(z)$ analytic in K except for possible poles in the points α_{nk} such that the first member of (62) is not greater than γ_2/γ_1.

We have several comments to make regarding the general significance of the problem studied in Theorems 19 and 20.

1°. Theorem 19 is a result concerning a problem which is invariant under arbitrary one-to-one conformal transformation of the region K, even under some circumstances (compare §11 below) if K is transformed into a region not bounded by a Jordan curve. But the full import of this invariant property appears best if the hypothesis (1) is replaced by the equivalent hypothesis (49).

2°. The two groups, Theorems 1–4 on the one hand and Theorems 19 and 20 on the other, are intimately related to each other, especially in view of Theorem 5. Theorem 1 can be employed as above to furnish a sequence $F_n(z)$ which converges as rapidly (in the sense indicated by (62)) as does any

sequence of admissible functions. Reciprocally, Theorem 20 shows that Theorem 19 can be employed to furnish a sequence of rational functions which converges maximally.

3°. Theorem 1 can be formulated so as to be invariant under arbitrary linear transformation of the complex variable. Condition (49) has the more general property of being essentially invariant under arbitrary one-to-one conformal transformation of the closed exterior of C_2. Remarks 1° and 2° serve to explain this latter invariant property, which has been in the past somewhat mystifying.

4°. Theorem 19 is a companion piece (Gegenstück) to results previously established by the present writer concerning interpolation by functions analytic and bounded in a given simply connected region K.* Prescribed poles on the one hand correspond to prescribed points of interpolation on the other, in each case satisfying asymptotic conditions in an annular region R interior to K, one of whose boundaries is the boundary of K. In the problem of interpolation there are two equivalent conditions on the prescribed points, completely analogous to (1) and (49). The function approximated to in the one problem and the function interpolated to in the other are assumed analytic in respective regions mutually complementary with respect to K, and convergence of approximating and interpolating functions is established throughout those (suitably chosen) regions of analyticity bounded by loci $\Phi = \text{const.}$ in R. Indeed, this entire pair of configurations exhibits a complete duality, in the sense already defined (op. cit., §8.3), except that the present duality is concerned with functions meromorphic or analytic in a region rather than with rational functions as in the former duality. Theorem 5 is sufficiently powerful to apply to both configurations.

11. **Generalizations and extensions.** In connection with the foregoing results, there suggests itself the question of extensions to regions of higher connectivity, and to regions not bounded by Jordan curves. We now proceed to discuss these questions in order.

We have hitherto limited ourselves to a region bounded merely by two Jordan curves C_1 and C_2 instead of two such sets of curves, entirely for the sake of simplicity and ease of exposition. Our methods are applicable with only minor modifications in the more general case. For instance we sketch rapidly the proof of

THEOREM 21. *Let R be a finite region bounded by two disjoint sets C_1 and C_2, each composed of a finite number of disjoint Jordan curves; let one of the Jordan curves C_1^0 of the former set contain in its interior all points of $C_1 - C_1^0 + C_2$. Let*

* These Transactions, vol. 46 (1939), pp. 46–65.

the points $\alpha_{n1}, \alpha_{n1}, \cdots, \alpha_{nn}$ *be separated from* R *by* C_1, *and let the points* $\beta_{n1}, \beta_{n2}, \cdots, \beta_{n,n+1}$ *not lying on* C_2 *be separated from* R *by* C_2. *Let* $\overline{R}$ *denote the closure of* R.

Let the relation (1) *hold uniformly on any closed set in* $\overline{R}$ *containing no point of* C_2. *Let the function* $\Phi(z)$ *be continuous in the closed region* $\overline{R}$, *and take constant values* γ_1 *and* $\gamma_2 < \gamma_1$ *on* C_1 *and* C_2 *respectively. Denote generically by* C_γ *the locus* $\Phi(z) = \gamma$ *in* $\overline{R}$, $\gamma_2 \leqq \gamma \leqq \gamma_1$. *For* $\gamma_2 < \gamma < \gamma_1$, *the locus* C_γ *consists of a finite number of mutually disjoint analytic Jordan curves, except that for each of a finite number of values of* γ *there are a finite number of multiple points, each common to a finite number of the Jordan curves composing* C_γ. *Denote generically by* R_γ *the open set composed of the interiors of the Jordan curves composing* C_2 *plus the point set in* $\overline{R}$ *on which we have* $\gamma_2 \leqq \Phi(z) < \gamma$; *thus* R_γ *need not be connected, but is the sum of a finite number of regions bounded by the entire locus* C_γ.

Let the function $f(z)$ *be single-valued and analytic on* R_γ *but not single-valued and analytic on any* $R_{\gamma'}$ *with* $\gamma' > \gamma$. *Then we have* (2), (3), *and* (4) *fulfilled, where* $r_n(z)$ *is the rational function of degree* n *whose poles lie in the points* α_{nk} *and which interpolates to* $f(z)$ *in the points* β_{nk}.

There exists no sequence of rational functions $R_n(z)$ *of respective degrees* n *with poles in the points* α_{nk} *such that either* (8) *or* (9) *is satisfied.*

Whenever a set of rational functions $R_n(z)$ *of respective degrees* n *with poles in the points* α_{nk} *satisfies* (20), *we shall say that the sequence* $R_n(z)$ *converges maximally to* $f(z)$ *on the set* C *composed of the Jordan curves* C_2 *and their respective interiors. Consequences of maximal convergence are equations* (22) *and* (23), *and if* Q *is an arbitrary continuum in* $\overline{R}$ *not a single point, also equations* (24)–(26).

In Theorem 21 the function $f(z)$ need not be a monogenic analytic function; nevertheless $f(z)$ can be extended analytically from the neighborhood of C_2 by paths in R_γ so as to be single-valued and analytic at every point of R_γ; but this property is shared by no set $R_{\gamma'}$ with $\gamma' > \gamma$. Then (a) for some path of analytic extension in R_γ from C_2 to C_γ the function $f(z)$ has a singularity on C_γ, or (b) the locus C_γ has a multiple point A and the analytic extensions of $f(z)$ from C_2 along paths in the various regions into which C_γ falls of which A is a boundary point fail to be analytic and identical throughout a neighborhood of A; or both (a) and (b) occur.

Theorem 5 extends to a region R bounded by two sets of Jordan curves, and the original proofs already given yield now Theorem 21.

With only minor modifications, which we now mention, the entire discussion of §§1–10 of the present paper is valid for the more general region R introduced in Theorem 21.

Under the hypothesis of Theorem 12 (that is to say, the hypothesis of Theorem 1), equation (44) may be written

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} G(z, \alpha_{nk}) = \log \Phi_1(z) \tag{74}$$

uniformly on any closed set in R, where

$$G(z, \alpha_{nk}) = \log \left| \frac{\bar{\phi}(\alpha_{nk})\phi(z) - 1}{\phi(z) - \phi(\alpha_{nk})} \right|$$

is Green's function* with pole in the point α_{nk} for the region exterior to the Jordan curves C_2. With this change in the manner of writing (44), the new formulation of Theorem 12 is obviously valid; theorem and proof extend without great difficulty to the situation of Theorem 21. But equation (44) is obviously simpler than (74), and more easily applied. Of course C_1 may consist of several Jordan curves even when C_2 is a single Jordan curve; under such circumstances (44) and similar relations need no revision.

Let us be more explicit about the extension of Theorem 12, using now the hypothesis of Theorem 21 rather than of Theorem 1; it is sufficient if (1) holds uniformly merely on any closed set in R. On an arbitrary closed set C' in R the functions $G(z, \alpha_{nk})$ are uniformly bounded, for the points α_{nk} are bounded from C', and the functions $G(z, \alpha_{nk})$ in R depend continuously on z and α_{nk} except in the neighborhood of α_{nk}. The functions

$$\frac{1}{n} \sum_{k=1}^{n} G(z, \alpha_{nk}) \tag{75}$$

are harmonic in R and uniformly limited on any closed set C' in R, hence form a normal family of harmonic functions in R. Let $\log \Phi_1(z)$ denote an arbitrary limit function of the family (75), the uniform limit on any closed set in R of a suitably chosen subsequence of the family. The functions (75) are continuous in the two-dimensional sense on C_2 and vanish there, so the subsequence converges uniformly on any closed set in $\overline{R}$ containing no point of C_1. The limit function $\log \Phi_1(z)$ is continuous in the two-dimensional sense on C_2 and vanishes there. Consider now the function

$$\frac{1}{n} \log \left| A_n \frac{(z - \beta_{n1}) \cdots (z - \beta_{n,n+1})}{(z - \alpha_{n1}) \cdots (z - \alpha_{nn})} \right| - \frac{1}{n} \log \sum_{k=1}^{n} G(z, \alpha_{nk}) - \frac{1}{n} \log G(z, \infty), \tag{76}$$

* The term Green's function is sometimes used for the negative of the present function.

which when suitably defined in the points α_{nk} is harmonic without exception in the region exterior to and bounded by the curves C_2. The sequence (76) or any subsequence converges uniformly on any closed set interior to R whenever that is true of the corresponding sequence or subsequence (75). By essentially the reasoning previously used in the proof of Theorem 12 we now have $\Phi(z)/\Phi_1(z) \equiv \gamma_2$ throughout R. Then the function $\Phi_1(z)$ is uniquely determined; from every subsequence of the sequence (75) can be extracted a new subsequence converging uniformly on any closed set in R to the function $\log \Phi_1(z)$, hence the sequence (75) itself converges uniformly on any closed set in R to the function $\log \Phi_1(z)$; the extension of Theorem 12 to the situation of Theorem 21 rather than of Theorem 1 is established.

In the extension of Theorem 13 to a region R bounded by more than two Jordan curves, we naturally replace (46) by

$$\log A_n = \sum_{k=1}^{n} G(\alpha_{nk}, \infty); \tag{77}$$

since only the modulus of A_n is of significance in (1), equation (77) is essentially equivalent to (46) under the hypothesis of Theorem 13. Of course the validity of (1) with the substitution (77) is to be established by the method indicated for the proof of (1) with the substitution (46) (op. cit., p. 274). The indicated extension of the remainder of §§1–10 now presents no difficulty, as the reader may verify; but in §8 a somewhat unusual convention is desirable regarding normal families of analytic functions (compare op. cit., §7.3).

We turn next to the question of replacing by more general sets the Jordan curves in Theorem 21 and in the corresponding results of §§1–10. Theorem 5 extends without difficulty and with only minor modifications to the case that R is bounded by a finite number of continua, none of which is a single point. There is no inherent difficulty in extending almost all of the discussion of the present paper to this more general case. However, with this more general hypothesis, such integrals as appear in (36) no longer have the meaning previously assigned, and points α_{nk} and β_{nk} cannot be taken or proved uniformly distributed on C_1 and C_2 by the previous methods. There are then two possible procedures for determining the points α_{nk} and β_{nk}: (i) consider integrals not over C_1 and C_2 but over sets of curves Γ_1 and Γ_2 in R and approaching C_1 and C_2 respectively, choosing points α_{nk} and β_{nk} on the sets Γ_1 and Γ_2 (as in op. cit., §4.4); but of course these points α_{nk} and β_{nk} then lie interior to R, and (1) may not be valid uniformly on C_1; (ii) if each point of C_1 and C_2 is a boundary point of the complement of $\overline{R}$, the harmonic function $U(x, y)$ in (36) may be uniformly approximated in $\overline{R}$ by functions harmonic throughout $\overline{R}$, by Lebesgue's theory of variable harmonic functions; these approximating

functions may be handled by the original method of uniform distribution of points, and yield points α_{nk} and β_{nk} exterior to $\overline{R}$ (compare op. cit., §4.4); this method can be used also for the appropriate points α_{nk} or β_{nk} if only one of the sets C_1 and C_2 has the property that each of its points is a boundary point of the complement of $\overline{R}$. It will be noticed that method (i) enables us to use for C, the point set on which the given function is approximated, any closed bounded set not separating the plane having only a finite number of components of which none is a single point. By an extension of Theorem 5 we may here take C an arbitrary closed bounded set whose complement is connected and possesses a Green's function.

Neither (i) nor (ii) enables us to establish uniform distribution on C_1 or C_2 of given points, and thus the necessity of the uniform distribution in the Corollary to Theorem 16, in Theorem 17, and in §9, is still unproved. In this connection there suggests itself another method: (iii) consider the integrals in a still more general sense, say in the sense of Stieltjes with respect to a variable defined on the boundary elements of the region; in the generalization of (36) this variable would be the conjugate of U, locally single-valued and continuous in $\overline{R}$. It seems to the writer probable that this method would be completely satisfactory both in determining points α_{nk} and β_{nk} from such an equation as (36), and in establishing the uniform distribution of given points as in the Corollary to Theorem 16; but this method has never been completely carried out.

We have hitherto considered merely a single region R, but it is possible (op. cit., §§8.7 and 8.8) to have such a relation as (1) valid simultaneously interior to two mutually disjoint regions of the kind considered in Theorem 1. Indeed if these regions R are preassigned, if the corresponding closed regions are mutually disjoint, and if no region separates another from the point at infinity, suitable harmonic functions can be used as generating functions to define the sequences α_{nk} and β_{nk} so that (1) is valid on any closed set in each region R. Then our entire discussion of §§1–10 carries over to the new situation, with a single exception. The function $\Phi(z)$ of (1) is harmonic in each R, and a locus C_γ occurs in each R. But it may occur that γ as defined from the function $f(z)$ is limited by the behavior of $f(z)$ in one region R, and not by the function $f(z)$ in a second region R; for instance $f(z)$ may be analytic throughout the latter. Theorem 5 then fails to apply to this second region R, and the conclusions (22)–(26) that we have drawn from Theorem 5 apply to the first region R but have not been established for the second region R.*

* But for a series of interpolation (78) the principal results are valid also for the second region R. For various other specific examples the corresponding statement is true; the general question is still open.

An important special case of Theorem 1 is that in which the α_{nk} and β_{nk} are independent of n. Under these circumstances the interpolating functions $r_n(z)$ of Theorem 1 are the partial sums of a series of interpolation

$$f(z) = a_0 + a_1\frac{z-\beta_1}{z-\alpha_1} + a_2\frac{(z-\beta_1)(z-\beta_2)}{(z-\alpha_1)(z-\alpha_2)} + \cdots, \tag{78}$$

and some of the results of Theorem 7 so far as it applies to the sequence $r_n(z)$ can be read off by elementary methods, without recourse to Theorem 5. For instance under the hypothesis of Theorem 1 we know (op. cit., §8.3) that the series in (78) converges to $f(z)$ uniformly on any closed set interior to C_γ, but by the definition of γ can converge uniformly throughout the closed interior of no $C_{\gamma'}$, $\gamma' > \gamma$. There follows (op. cit., §3.4, Theorem 5) the relation

$$\limsup_{n\to\infty} \left| a_n/A_n \right|^{1/n} = 1/\gamma. \tag{79}$$

For the series (78) we deduce by inspection inequalities (2)–(4). Impossibility of the inequality sign in (2) and (4) follows at once from (1) and (79). Much of Theorem 7 for the series (78) is also a consequence of (1) and (79).

In generalizing the results of the present paper, it is also of interest to notice that points α_{nk} and β_{nk} interior to R may be admitted in restricted numbers, and that the equation (1) need not be supposed to hold uniformly on any closed set in R; under suitable circumstances it is sufficient if (1) holds uniformly on C_1 and on each C_λ for λ everywhere dense in the interval $\gamma_2 < \lambda < \gamma_1$. A sufficiently general instance of this remark has already been elaborated elsewhere (op. cit., §8.5).

HARVARD UNIVERSITY,
CAMBRIDGE, MASS.

NOTE ON THE DEGREE OF CONVERGENCE OF SEQUENCES OF ANALYTIC FUNCTIONS*

BY

J. L. WALSH

It is the object of this note first to present a general theorem (Theorem 1) on maxima and superior limits of sequences of analytic functions; this is essentially a result in the theory of subharmonic functions, presumably of some interest in itself. Second, Theorem 1 is employed to establish a sharper theorem (Theorem 2) on the precise degree of convergence of general sequences of analytic functions. Finally, these general results are used to establish a quite specific result (Theorem 3) on the degree of convergence of certain sequences of analytic functions in an annular or more general region. Parts of Theorem 3 have previously been established, but in a more restricted form and by another method. The present Theorem 3 is sufficiently general to dominate results established elsewhere by the author on maximal convergence of sequences of polynomials,† on approximation by functions analytic in a given region,‡ on interpolation by functions analytic in a given region,§ and on maximal convergence of sequences of rational functions.‖ Various other applications of Theorem 3 seem to suggest themselves and will presumably appear in the future.¶

The following treatment is chosen rather for elements of simplicity than for ultimate generality. Some extensions present themselves at once as obvious; others are scarcely less immediate. But even in its present form, Theorem

* Presented to the Society, December 30, 1938, under the title *Maximal convergence of sequences of rational functions*; received by the editors October 2, 1939.

† *Interpolation and Approximation by Rational Functions in the Complex Domain*, American Mathematical Society Colloquium Publications, vol. 20, New York, 1935. All references in the present note not otherwise indicated are to this book.

‡ Proceedings of the National Academy of Sciences, vol. 24 (1938), pp. 477–486.

§ These Transactions, vol. 46 (1939), pp. 46–65.

‖ These Transactions, vol. 47 (1940), pp. 254–292.

¶ In particular the writer hopes later to apply these methods to the study of such questions as overconvergence, gap theorems, and the analogues of Jentzsch's theorem.

No general study of these questions seems as yet to have been made for sequences $f_n(z)$ of the type considered in Theorem 3, as contrasted with the partial sums of a series $\sum a_n p_n(z)$ when the asymptotic character of the $p_n(z)$ is known. But certain results on this general study for the special case of maximal sequences of *polynomials* have recently been obtained by Dr. Z. I. Mosesson in his 1937 doctoral dissertation at Harvard University. Mosesson's results are still unpublished, but were obtained prior to those of the present note; for the special case of sequences of polynomials, they overlap slightly our Theorem 3. Mosesson employs a new form of the two-constant theorem.

3 seems to be of sufficient power and importance to deserve a systematic and independent exposition.

The following theorem is largely an application of known principles; nevertheless the present form is highly convenient for reference:

THEOREM 1. *Let R be a region whose boundary B consists of a finite number of Jordan arcs J_k, mutually disjoint except possibly at end points A_j. Let the function $V(z)$ be harmonic and bounded in R, continuous in the corresponding closed region $\overline{R}$ except in the points A_j, constant on each J_k.*

Let the functions $F_n(z)$, $n=1, 2, \cdots$, be uniformly bounded in R, locally analytic except perhaps for branch points although not necessarily single-valued in R, but let $|F_n(z)|$ be single-valued in R and continuous in $\overline{R}$ except perhaps in the points A_j; and suppose we have for each k

$$\limsup_{n\to\infty} [\text{l.u.b. } |F_n(z)|, z \text{ on } J_k] \leqq [e^{V(z)}, z \text{ on } J_k]. \tag{1}$$

Then on every closed set Q in $\overline{R}$ containing no point A_j we have

$$\limsup_{n\to\infty} [\max |F_n(z)|, z \text{ on } Q] \leqq [\max e^{V(z)}, z \text{ on } Q]. \tag{2}$$

In the proof of Theorem 1 let us suppose for the moment that $V(z)$ and $|F_n(z)|$ are continuous in $\overline{R}$. If $\eta>0$ is arbitrary, we have from (1) for n sufficiently large

$$\log |F_n(z)| \leqq V(z) + \eta, \qquad z \text{ on } B. \tag{3}$$

To be sure, the function $F_n(z)$ may vanish in $\overline{R}$, so here and below the value $-\infty$ is contemplated as admissible in the first member of (3). The subharmonic character of the function $\log |F_n(z)|$ in R now yields from (3)

$$\log |F_n(z)| \leqq V(z) + \eta, \qquad z \text{ in } \overline{R}, \tag{4}$$

for n sufficiently large, and (2) therefore follows.

If $V(z)$ and $|F_n(z)|$ are no longer assumed continuous in the points A_j, the reasoning just given is to be modified slightly. It is no loss of generality to assume, as we do, that R is a finite region bounded by analytic Jordan curves; for this situation can be obtained by transforming the original configuration by a suitable conformal map. Let $\eta>0$ be arbitrary. Denote by M a common upper bound for $\log |F_n(z)|$ and $V(z)$ in R, and denote by $W_\nu(z)$ a function which is harmonic in R, continuous in $\overline{R}$, equal to $2M+\eta$ in each point A_j, never greater than $2M+\eta$ on B, and equal to zero on B except in the neighborhood of diameter ν of each point A_j. When the diameters ν of these neighborhoods approach zero, the function $W_\nu(z)$ approaches zero at every interior point of R, as follows from the representation of $W_\nu(z)$ in R

by Green's formula involving Green's function for R. For n sufficiently large we have from (1)

$$[\text{l.u.b. } \log |F_n(z)|, z \text{ on } J_k] \leqq [V(z) + \eta, z \text{ on } J_k].$$

The function $V_n'(z) \equiv \log |F_n(z)| - V(z) - \eta$ is subharmonic in R. When z in R approaches B, the function $V_n'(z) - W_\nu(z)$ has a superior limit not greater than zero, whether z approaches a point A_j or an interior point of J_k; it follows that the superior limit of this function in R is not greater than zero:

$$V_n'(z) \leqq W_\nu(z), \qquad z \text{ in } R.$$

This inequality is true for each ν, so we have by allowing ν to approach zero,

$$V_n'(z) \leqq 0, \qquad z \text{ in } R,$$

which is equivalent to the inequality (4), valid for n sufficiently large at every point of R; by allowing z in R to approach B, we see that inequality (4) holds throughout $\overline{R}$, except in the points A_j where $V(z)$ is not continuous. Inequality (2) follows, so Theorem 1 is completely established.

Theorem 1 persists if the functions $F_n(z)$ are not required to be continuous on the arcs J_k except in the points A_j, provided (1) is valid, where the square bracket in the first member of (1) is interpreted as the least upper bound of the set of limit values approached by $|F_n(z)|$ when z in R approaches an arbitrary point of J_k, end points of J_k excepted. Similarly, if Q contains points of B, the square bracket in the first member of (2) is to be interpreted as the least upper bound of the set of values approached by $|F_n(z)|$ as z in R approaches an arbitrary point of Q. The proof just given is valid in this more general case.

In a similar manner we may interpret and establish (2) even when the $F_n(z)$ are not required to be continuous on the arcs J_k except in the points A_j and when Q in $\overline{R}$ is allowed to contain points A_j.

A result complementary to Theorem 1 is now to be proved:

THEOREM 2. *Under the conditions of Theorem* 1 *suppose the equality sign holds in* (2) *for a particular continuum* Q_0 *in* R (Q_0 *may be a single point*):

$$\limsup_{n\to\infty} [\max |F_n(z)|, z \text{ on } Q_0] = [\max e^{V(z)}, z \text{ on } Q_0]. \tag{5}$$

Then on every continuum Q *in* $\overline{R}$ *consisting of more than one point and containing no point* A_j *we have the corresponding equality*:

$$\limsup_{n\to\infty} [\max |F_n(z)|, z \text{ on } Q] = [\max e^{V(z)}, z \text{ on } Q]. \tag{6}$$

Our proof of Theorem 2 is indirect. Let us suppose first that Q lies in R. If (6) is not valid we may assume

$$\limsup_{n\to\infty} [\max\ |F_n(z)|,\ z \text{ on } Q] = M, \tag{7}$$

$$M < [\max\ e^{V(z)},\ z \text{ on } Q]. \tag{8}$$

We are to show the impossibility of (7) and (8). Replacing Q by a proper subset of Q (still a continuum not a single point) without altering the second member of (6) leaves us with a stronger statement to be proved and hence is allowable. Thus it is no loss of generality, since $V(z)$ is continuous in $\overline{R}$ except in the points A_j, to assume that we have

$$V(z) > \log M, \qquad z \text{ on } Q. \tag{9}$$

We assume also, as we may do, that the complement of Q is simply connected.

Denote by $V_1(z)$ the function harmonic and bounded in the region R_1: $R-Q$, continuous and equal to $V(z)$ on each J_k except in the points A_j, and continuous and equal to $\log M$ on the boundary points of R_1 which belong to Q. Theorem 1 in generalized form applies to the region R_1; for the conformal map $z=\Psi(w)$ of the complement of Q onto the interior of a circle C maps R onto a region satisfying the original hypothesis of Theorem 1. On the circle C we have by (7)

$$\limsup_{n\to\infty} [\max\ |F_n[\Psi(w)]|,\ w \text{ on } C] \leqq M, \tag{10}$$

provided the square bracket in the first member of (10) is interpreted as the least upper bound of all limit values of $|F_n[\Psi(w)]|$ as w approaches C. For the region R_1 no point A_j lies on Q. From Theorem 1 follows by (2) the inequality

$$\limsup_{n\to\infty} [\max\ |F_n(z)|,\ z \text{ on } Q_0\cdot\overline{R}_1] \leqq [\max\ e^{V_1(z)},\ z \text{ on } Q_0\cdot\overline{R}_1];$$

a glance at the proof of Theorem 1 shows that the possible lack of one-to-one-ness of the transformation $z=\Psi(w)$ on Q and C plays no role; the relation (4) valid on an open set in R is valid also on the corresponding closed set. By virtue of (7) and (8) the corresponding inequality holds also in points of $Q_0\cdot Q$, provided we define $V_1(z)$ as equal to $\log M$ in the interior points of Q:

$$\limsup_{n\to\infty} [\max\ |F_n(z)|,\ z \text{ on } Q_0\cdot Q] \leqq [\max\ e^{V_1(z)},\ z \text{ on } Q_0\cdot Q].$$

Combination of these two inequalities gives us

$$\limsup_{n\to\infty} [\max\ |F_n(z)|,\ z \text{ on } Q_0] \leqq [\max\ e^{V_1(z)},\ z \text{ on } Q_0]. \tag{11}$$

However, the function $V(z)-V_1(z)$ is harmonic and bounded in R_1, continuous in the corresponding closed region except in the points A_j, zero on the J_k, and by (9) is positive in the boundary points of R_1 which belong to Q. Consequently we have throughout R_1

$$V(z) > V_1(z). \tag{12}$$

Inequality (12) is also valid in the points of Q, hence throughout R. The functions $V(z)$ and $V_1(z)$ are continuous on the closed set Q_0 in R, so we have

$$[\max V(z),\ z \text{ on } Q_0] > [\max V_1(z),\ z \text{ on } Q_0].$$

Thus inequality (11) contradicts (5), and thereby equation (6) is established provided Q lies in R.

The proof just given may require modification if Q intersects B, as we now suppose. If Q is identical with $\overline{R}$, the proof is immediate, for no point A_j can exist and the function $V(z)$ is identically constant on each component of B; the negation of (6) yields a majorant $V_1(z)$ defined throughout $\overline{R}$ and which may replace $V(z)$ in Theorem 1. We require that $V_1(z)$ shall be harmonic in R, continuous in $\overline{R}$, equal to $\log M$ as defined by (7) on the component or components of B on which $V(z)$ takes its maximum value in $\overline{R}$, and equal to $V(z)$ on the remaining components of B. In R we have (12) as before. It follows from Theorem 1 that when Q_0 lies in R the first member of (5) is not greater than $[\max e^{V_1(z)},\ z \text{ on } Q_0]$, which is less than the second member of (5); this contradicts (5) and thereby establishes (6).

If Q in $\overline{R}$ is not identical with $\overline{R}$ but intersects B, the set Q separates R into one or more subregions each containing no point of Q. Denote by R_1 an arbitrary one of these subregions. At least one boundary point of R_1 is a point of Q. Of course we suppose (7), (8), and (9) valid on Q as before; we are to reach a contradiction.

We assume as before that the complement of Q is simply connected, and assume also that Q intersects but one component of B. Let B_1 be the boundary of R_1. We denote by $V_1(z)$ the function bounded and harmonic in R_1, continuous and equal to $\log M$ in the points of B_1 which belong to Q, and continuous and equal to $V(z)$ in the points of B_1 not points A_j and not belonging to Q. Whether or not B_1 lies entirely in B, some component of B_1 contains some point D of B_1 such that every point of B_1 in a neighborhood of D belongs to Q. Map onto the interior of a circle C the complement of that component of B_1 (there can be but one such component) which contains points of Q. In this conformal map the boundary elements (prime ends) of R_1 which consist of points of Q correspond to precisely an entire arc of C, possibly the entire circumference. From this map it becomes clear that Theorem 1 in generalized

form applies to the region R_1, and that (12) is valid in R_1. Consequently we have (compare (11))

$$\limsup_{n\to\infty} [\max \mid F_n(z) \mid, z \text{ on } Q_1] \leqq [\max e^{V_1(z)}, z \text{ on } Q_1], \tag{13}$$

provided merely that Q_1 is an arbitrary continuum which lies in R_1. Then Q_1 also lies in R, and inequalities (12) and (13) contradict (6) as already established for a continuum which lies in R. Theorem 2 has now been completely proved.

It is to be noted that under the conditions of Theorem 2 the conclusion (6) may fail if we do not suppose Q to consist of more than one point. This is illustrated by the example R: $|z|<1$, $F_n(z)\equiv z^{1/n}$, $V(z)\equiv 0$. On every continuum Q in $\overline{R}$ consisting of more than one point the first member of (6) has the value unity; but if Q reduces to the single point $z=0$, the first member of (6) is zero.

We shall now apply Theorems 1 and 2 to a configuration in which the function $V(z)$ takes on only two distinct values on B, and in which the loci $V(z)=$ const. in R have especially simple topological properties. This configuration occurs rather frequently in various studies on interpolation and approximation (loc. cit.).

THEOREM 3. *Let R be a region bounded by a finite number of mutually disjoint Jordan curves K_0 and a finite number of mutually disjoint Jordan curves K_{-1} disjoint with K_0. Let the function $U(z)$ be harmonic in R, continuous in the corresponding closed region $\overline{R}$, equal to zero and -1 on K_0 and K_{-1} respectively. Denote generically by K_σ the locus $U(z)=\sigma$, $0>\sigma>-1$, in R, by R_σ the open set $\sigma>U(z)>-1$ in R bounded by K_σ and K_{-1}; then R_σ is a sum of regions; the closure of R_σ is denoted by $\overline{R}_\sigma$.*

Let the function $f(z)$ be analytic throughout R_ρ but (considered with its possible analytic extensions) not be analytic throughout any $R_{\rho'}$ with $0>\rho'>\rho$, and let $f(z)$ be continuous in the two-dimensional sense on K_{-1} with respect to the domain $\overline{R}$. Let the functions $f_n(z)$ for $n=1, 2, \cdots$ be analytic in R, continuous in $\overline{R}$, with

$$\limsup_{n\to\infty} [\max \mid f_n(z) \mid, z \text{ on } K_0]^{1/n} \leqq e^\alpha > 1, \tag{14}$$

$$\limsup_{n\to\infty} [\max \mid f(z) - f_n(z) \mid, z \text{ on } K_{-1}]^{1/n} \leqq e^\beta < 1. \tag{15}$$

Then we must have

$$\alpha + \alpha\rho - \beta\rho \geqq 0: \tag{16}$$

moreover we have

$$\limsup_{n\to\infty} [\max \mid f_n(z) \mid, z \text{ on } K_\mu]^{1/n} \leqq e^{\alpha+\alpha\mu-\beta\mu}, \tag{17}$$

provided $\alpha+\alpha\mu-\beta\mu \geqq 0$ (*this inequality is satisfied if we have* $0 \geqq \mu \geqq \rho$), *and we have*

$$\limsup_{n\to\infty} [\max \mid f(z) - f_n(z) \mid, z \text{ on } K_\sigma]^{1/n} \leqq e^{\alpha+\alpha\sigma-\beta\sigma}, \tag{18}$$

provided $\alpha+\alpha\sigma-\beta\sigma<0$ (*this inequality is satisfied for* σ *sufficiently near* -1).

If the equality sign holds in (16):

$$\alpha + \alpha\rho - \beta\rho = 0, \tag{16'}$$

then the equality sign holds also in (17) *and* (18):

$$\limsup_{n\to\infty} [\max \mid f_n(z) \mid, z \text{ on } K_\mu]^{1/n} = e^{(\alpha-\beta)(\mu-\rho)}, \qquad 0 \geqq \mu \geqq \rho, \tag{17'}$$

$$\limsup_{n\to\infty} [\max \mid f(z) - f_n(z) \mid, z \text{ on } K_\sigma]^{1/n} = e^{(\alpha-\beta)(\sigma-\rho)}, \qquad \rho > \sigma \geqq -1. \tag{18'}$$

Indeed, if Q_σ *is a continuum in* $\overline{R}$ *consisting of more than one point and lying in* $\overline{R}_\sigma$ *but having at least one point in common with* K_σ, *then we have*

$$\limsup_{n\to\infty} [\max \mid f_{n+1}(z) - f_n(z) \mid, z \text{ on } Q_\sigma]^{1/n} = e^{(\alpha-\beta)(\sigma-\rho)}, \tag{19}$$

and we have

$$\limsup_{n\to\infty} [\max \mid f(z) - f_n(z) \mid, z \text{ on } Q_\sigma]^{1/n} = e^{(\alpha-\beta)(\sigma-\rho)} \qquad \text{if } \sigma < \rho, \tag{20}$$

$$\limsup_{n\to\infty} [\max \mid f_n(z) \mid, z \text{ on } Q_\sigma]^{1/n} = e^{(\alpha-\beta)(\sigma-\rho)} \qquad \text{if } \sigma \geqq \rho. \tag{21}$$

From the general properties of harmonic functions it follows (as in op. cit., §§3.3 and 4.1) that each K_σ in R consists of a finite number of disjoint analytic Jordan curves, except that for each of a finite number of values of σ the curves may have a finite number of multiple points. At such a multiple point the tangents to the various branches of K_σ are equally spaced. Each K_σ separates each point of K_0 from each point of K_{-1}. The open set R_σ is not necessarily connected, but consists of one or more regions, each bounded by points of K_σ and points of K_{-1}.

It is not the intention in Theorem 3 to demand that $f(z)$ shall be a monogenic analytic function. The function $f(z)$ shall be single-valued and analytic at every point of R_ρ, but there shall be no function single-valued and analytic throughout the interior of any $R_{\rho'}$, $\rho'>\rho$, which coincides with $f(z)$ in R_ρ. Thus (a) some branch of $f(z)$ found by analytic extension in R_ρ from the neighborhood of K_{-1} has a singularity on K_ρ; or (b) the locus K_ρ in R has a

multiple point, and the monogenic functions defined at that multiple point by analytic extension in the various separated parts of R_ρ from the neighborhood of K_{-1} are not identical; or (c) both (a) and (b) occur.

Of course $\alpha+\alpha\sigma-\beta\sigma$ is merely a linear function of σ, which for $\sigma=0$ reduces to α and for $\sigma=-1$ reduces to β.*

From inequality (14) we may write for arbitrary $\epsilon>0$ and for n sufficiently large

$$|f_n(z)| \leqq e^{n(\alpha+\epsilon)}, \qquad |f_{n+1}(z)| \leqq e^{n(\alpha+\epsilon)}, \qquad z \text{ on } K_0,$$

$$|f_{n+1}(z) - f_n(z)| \leqq 2e^{n(\alpha+\epsilon)}, \qquad z \text{ on } K_0,$$

$$\limsup_{n\to\infty} [\max |f_{n+1}(z) - f_n(z)|, z \text{ on } K_0]^{1/n} \leqq e^{\alpha}. \tag{22}$$

From inequality (15) we may similarly write

$$\limsup_{n\to\infty} [\max |f_{n+1}(z) - f_n(z)|, z \text{ on } K_{-1}]^{1/n} \leqq e^{\beta}. \tag{23}$$

We are now in a position to apply Theorems 1 and 2, identifying the region R of Theorem 1 with the region R of Theorem 3, setting

$$F_n(z) \equiv [f_{n+1}(z) - f_n(z)]^{1/n},$$

and setting

$$V(z) \equiv \alpha + (\alpha - \beta)U(z).$$

Inequality (2) yields at once $(0>\sigma>-1)$

$$\limsup_{n\to\infty} [\max |f_{n+1}(z) - f_n(z)|, z \text{ on } K_\sigma]^{1/n} \leqq e^{\alpha+\alpha\sigma-\beta\sigma}. \tag{24}$$

We are now in a position to establish (16). It follows from (15) that $f_n(z)$ approaches $f(z)$ uniformly on K_{-1}. If (16) is not true, we have $\alpha+\alpha\rho-\beta\rho<0$, from which it follows that we have $\alpha+\alpha\rho_1-\beta\rho_1<0$ for a suitably chosen ρ_1, $0>\rho_1>\rho$. Then by (24) for $\sigma=\rho_1$, the sequence $f_n(z)$ converges uniformly on K_{ρ_1}, hence converges uniformly in the closed set $\overline{R}_{\rho_1}$, to some function which we denote by $F(z)$, analytic in R_{ρ_1} and continuous in $\overline{R}_{\rho_1}$. The function $F(z)-f(z)$ is analytic in R_ρ, continuous on K_{-1} with respect to the region R, and vanishes on K_{-1}. Consequently (see for instance op. cit., §1.9) the function $F(z)-f(z)$ vanishes identically in the neighborhood of K_{-1} interior to R, and vanishes identically whenever defined by analytic extension from that neighborhood. The function $F(z)$ is single-valued and analytic throughout

* It may seem to the reader somewhat clumsy to employ the values zero and -1 for $U(z)$ on the two parts of the boundary of R. Such choice is, however, rather convenient in the applications of Theorem 3. If any other two constants are used instead, conditions (16) and (16′) are easily transformed into the new notation by means of the linearity property.

R_{ρ_1}; hence $f(z)$ can be defined so as to be single-valued and analytic throughout R_{ρ_1}, contrary to hypothesis. This contradiction completes the proof of (16).

Inequality (24) implies for arbitrary $\epsilon>0$ and for n sufficiently large

$$|f_{n+1}(z)-f_n(z)| \leqq e^{(\alpha+\alpha\sigma-\beta\sigma+\epsilon)n}, \qquad z \text{ on } K_\sigma, \tag{25}$$

so merely by writing

$$\begin{aligned} f_{n+1}(z) \equiv f_N(z) &+ [f_{N+1}(z)-f_N(z)] + [f_{N+2}(z)-f_{N+1}(z)] \\ &+ \cdots + [f_{n+1}(z)-f_n(z)], \qquad n \geqq N, \end{aligned} \tag{26}$$

we deduce from (25) inequality (17), for all values of μ for which $\alpha+\alpha\mu-\beta\mu$ is nonnegative, hence certainly for all μ satisfying the inequality $0 \geqq \mu \geqq \rho$.

If we have $\alpha+\alpha\sigma-\beta\sigma<0$, and if $\epsilon>0$ is chosen so small that also $\alpha+\alpha\sigma-\beta\sigma+\epsilon<0$, then by (25) the sequence $f_n(z)$ converges uniformly on K_σ, hence by (15) converges uniformly in $\overline{R}_\sigma$ to some function $F(z)$ which by the reasoning used above is identical with $f(z)$ in $\overline{R}_\sigma$. Inequality (25) holds for n sufficiently large, and the identity

$$f(z)-f_n(z) \equiv [f_{n+1}(z)-f_n(z)] + [f_{n+2}(z)-f_{n+1}(z)] + \cdots \tag{27}$$

then gives us inequality (18).

If the equality sign holds in (16), that is to say if (16′) is valid, we proceed to establish

$$\limsup_{n\to\infty} [\max |f_{n+1}(z)-f_n(z)|, z \text{ on } K_\rho]^{1/n} = 1. \tag{28}$$

If (28) does not hold, it follows from (25) that we have

$$\limsup_{n\to\infty} [\max |f_{n+1}(z)-f_n(z)|, z \text{ on } K_\rho]^{1/n} \leqq e^{\gamma} < 1; \tag{29}$$

we show that (29) leads to a contradiction. Apply Theorem 1 to the region or regions in R between and bounded by K_0 and K_ρ, setting as before

$$F_n(z) \equiv [f_{n+1}(z)-f_n(z)]^{1/n},$$

and now with the choice

$$V(z) \equiv \alpha + (\gamma-\alpha)U(z)/\rho.$$

By virtue of (22) and (29), Theorem 1 applies here. On K_ρ we have $V(z)=\gamma<0$, so on a suitably chosen K_{ρ_1} with $0>\rho_1>\rho$ we also have $V(z)<0$. The corresponding inequality (2) for each component of K_{ρ_1} implies uniform

convergence of the sequence $f_n(z)$ on K_{ρ_1}, which we have already shown to be impossible. This contradiction establishes equation (28).

Theorem 2 now applies to the original region R, with $V(z) \equiv \alpha + (\alpha - \beta)U(z)$, thanks to (28) as a consequence of (16′), and establishes from (22) and (23) the equation

$$(30) \quad \limsup_{n\to\infty} [\max \mid f_{n+1}(z) - f_n(z) \mid, z \text{ on } K_\sigma]^{1/n} = e^{(\alpha-\beta)(\sigma-\rho)}, \quad 0 \geqq \sigma \geqq -1.$$

By use of (26) and (27) respectively, equation (30) implies

$$(31) \quad \limsup_{n\to\infty} [\max \mid f_n(z) \mid, z \text{ on } K_\mu]^{1/n} \leqq e^{(\alpha-\beta)(\mu-\rho)}, \qquad 0 \geqq \mu \geqq \rho,$$

$$(32) \quad \limsup_{n\to\infty} [\max \mid f(z) - f_n(z) \mid, z \text{ on } K_\sigma]^{1/n} \leqq e^{(\alpha-\beta)(\sigma-\rho)}, \qquad \rho > \sigma \geqq -1.$$

On the other hand, elementary inequalities show that the inequality sign in (31) or (32) would contradict (30). Consequently equations (17′) and (18′) are established.

Equation (19) follows from equation (28) and Theorem 2 applied to the original region R by use of (22) and (23). Equations (20) and (21) follow from (19) precisely as we have derived (17′) and (18′) from (30). Theorem 3 is completely proved.

In connection with Theorem 3 it may be noticed that inequalities (17) and (18) may both be expressed in essence in the form

$$(33) \quad \limsup_{n\to\infty} [\max \mid f_{n+1}(z) - f_n(z) \mid, z \text{ on } K_\sigma]^{1/n} \leqq e^{\alpha+\alpha\sigma-\beta\sigma},$$

with no restriction that $\alpha+\alpha\sigma-\beta\sigma$ should be positive, negative, or zero. But (18) is more specific than (33), in the sense that (18) exhibits the limit $f(z)$ of the sequence $f_n(z)$ if the second member of (18) is less than unity. Similarly, inequalities (17′) and (18′) may be combined in the single form (30), which is true for every σ, $0 \geqq \sigma \geqq -1$, provided (16′) is valid. We formulate

COROLLARY 1. *Under the hypothesis of Theorem 3, inequality* (33) *is valid for every* σ, $0 \geqq \sigma \geqq -1$; *if equation* (16′) *holds, then equation* (30) *is valid for every* σ, $0 \geqq \sigma \geqq -1$.

In the nature of a converse to Corollary 1 we have

COROLLARY 2. *Let R and $f(z)$ satisfy the hypothesis of Theorem 3. Let the sequence of functions $f_n(z)$ analytic in R and continuous in $\overline{R}$ converge uniformly to $f(z)$ on K_{-1}, and satisfy the inequalities* (22) *and* (23), *with $\alpha > 0$, $\beta < 0$. Then the conclusions of Theorem 3 and of Corollary 1 are fulfilled.*

The proof of Corollary 2 is contained in our previous discussion, for we

have used essentially (22) and (23) rather than (14) and (15) in the proof of Theorem 3.

For the sake of simplicity we have assumed in Theorem 3 and its corollaries that R is bounded by a finite number of mutually disjoint Jordan curves K_0 and a finite number of mutually disjoint Jordan curves K_{-1} disjoint with K_0; and have assumed $f(z)$ to be continuous on K_{-1} in the two-dimensional sense with respect to $\overline{R}$, and $f_n(z)$ to be continuous in $\overline{R}$. *Theorem 3 and its corollaries persist if these requirements are somewhat weakened, namely replaced by the requirement that R be bounded by two disjoint sets K_0 and K_{-1}, each of which consists of a finite number of components none of which is a single point; and the requirement that $f(z)$ is bounded in the neighborhood of K_{-1} in R, and that $f_n(z)$ is bounded in R.* Under these new conditions on R it is no loss of generality to assume that R fulfills the original conditions, for the method of successive conformal mapping onto the interior of a circle of simply-connected regions containing R each bounded by a single component of the boundary of R maps R onto a region fulfilling the original requirements on R and which is even bounded by a finite number of analytic Jordan curves. The function $U(z)$ is invariant under such transformation. In such expressions as appear in (14), (15), (22), (23), the term "max" is now to be interpreted as the least upper bound of the limit values as z in R approaches K_0 or K_{-1}. We may also take here not least upper bound of all limit values, but least upper bound of boundary values for normal approach in the sense of Fatou; these values exist almost everywhere (we assume R bounded by analytic Jordan curves), and may be used also in Theorem 1; the least upper bound of limit values cannot exceed the least upper bound of Fatou boundary values. With the new hypothesis and this new interpretation of notation, the previous reasoning remains essentially valid.*

It is fairly obvious that Theorem 3 extends to the case that K_0 and K_{-1} may have a finite number of points in common. But here the topological nature of the loci K_σ may become rather complicated. Indeed, Theorem 3 extends also to the case that $U(z)$ is no longer piecewise constant on the boundary B of R, but is bounded in R and continuous in $\overline{R}$ except at a finite number of points. Under these conditions we replace inequalities (22) and (23) of Corollary 2 by such a relation as

* We need, however, some such proposition as the following: *If the function $\phi(z)$ is analytic and bounded for $r \leqq |z| < 1$, and if the boundary values of $\phi(z)$ for normal approach to $|z| = 1$ vanish almost everywhere on $|z| = 1$, then $\phi(z)$ vanishes identically.* Form Cauchy's integral for z in the annular region $r \leqq |z| \leqq r_1 < 1$, the sum of an integral over $|z| = r$ and an integral over $|z| = r_1$; as r_1 approaches unity the Cauchy integral for $\phi(z)$ over $|z| = r_1$ approaches zero and therefore vanishes identically; hence $\phi(z)$ can be extended analytically across $|z| = 1$ so as to be analytic for every z with $|z| \geqq r$. Since $\phi(z)$ has the value zero almost everywhere on the circle $|z| = 1$, that function vanishes identically.

$$\limsup_{n\to\infty} \{\text{l.u.b.}\ [e^{V(z)} - |f_{n+1}(z) - f_n(z)|^{1/n}], z \text{ on } B\} \geqq 0,$$

where as before the square bracket refers to limiting values on B. Theorems 1 and 2 admit a corresponding extension, where $V(z)$ is merely harmonic and bounded in R, and continuous in $\overline{R}$ except at a finite number of points.

We establish still another corollary to Theorem 3:

COROLLARY 3. *Under the conditions of Theorem* 3 *with* (16′) *fulfilled, the sequence* $f_n(z)$ *converges throughout no region lying in* R *but not in* R_ρ.

If the sequence $f_n(z)$ converges in a region which lies in R but not in R_ρ, that sequence converges uniformly in a region Q which lies in R but not in R_ρ, for it follows from a theorem due to Osgood that every region of convergence contains a region of uniform convergence. For the particular region Q we have

$$\limsup_{n\to\infty} [\max\ |f_{n+1}(z) - f_n(z)|, z \text{ on } Q]^{1/n} \leqq 1,$$

in contradiction to (19).

The two-constant theorem can be employed with ease in the proof of Theorem 3. Indeed, the conclusions (16), (17), (18), (17′), (18′) can be established by the method used for the preliminary results (in the papers already mentioned) which suggested Theorem 3 and are contained in it. Also the conclusions (19), (20), (21) can be proved by the two-constant theorem provided $\sigma > -1$, by an extension of that same method. These proofs are less immediate than the present ones, namely applications of Theorem 2. The writer doubts that (19), (20), (21) can be proved by the two-constant theorem if $\sigma = -1$, that is to say, if Q_σ is an arc of K_{-1}.

HARVARD UNIVERSITY,
CAMBRIDGE, MASS.

ON A DECOMPOSITION OF TRUE CYCLIC ELEMENTS*

BY

DICK WICK HALL†

INTRODUCTION
PART I. DEVELOPMENT OF THE THEORY
PART II. APPLICATIONS
PART III. EXAMPLES

INTRODUCTION

The study of locally connected continua has been greatly simplified by the introduction of the cyclic element concept of G. T. Whyburn.‡ By definition, a cyclic element of a locally connected continuum M is any subset E of M satisfying some one of the following conditions: (1) E is a cut point of M; (2) E is an end point of M; (3) E is any nondegenerate subcontinuum of M containing no cut point of itself which is saturated§ in M with respect to this property. Every cyclic element of type (3) is called a true cyclic element.

Along these lines much light has been shed upon the knotty problem of the structure of a locally connected continuum M by considering M as the sum of its cyclic elements and then deducing properties of M from the properties of cyclic elements and the known relationships which exist between them.

The richness of the results in this direction has led to the hope that some still finer decomposition of locally connected continua may be possible. One such finer decomposition, based on combinatorial concepts, is given by the cyclic elements of higher order.‖ It is to the problem of obtaining a decomposition in the set theoretic direction that the present paper is devoted. We begin with a true cyclic element C as our space and attempt to decompose it into subsets having desirable properties. The decomposition of the space into its cyclic elements then gives the decomposition we want.

This paper is but a first step towards the ultimate goal of an extensive structure theory of true cyclic elements. It explores only the most fundamen-

* Presented to the Society, April 9, 1938, and October 29, 1938; received by the editors June 26, 1939, and, in revised form, December 19, 1939.

† This paper was completed while the author was a National Research Fellow at the University of Pennsylvania.

‡ See Kuratowski and Whyburn, Fundamenta Mathematicae, vol. 16 (1930), pp. 305–331. In this article an extensive bibliography may be found.

§ A subset E of M is said to be saturated in M with respect to a property P provided E has property P and no subset of M containing E but distinct from E has this property.

‖ See G. T. Whyburn, American Journal of Mathematics, vol. 56 (1934), pp. 133–146.

tal properties of our new elements, and the most elementary relationships existing between pairs of such sets and between such sets and their complements. It leaves for future study such considerations as the following:

(1) The possible existence of analogues of simple cyclic chains.

(2) The possible existence of analogues of A-sets.

(3) The study of the hyperspace formed when each secondary element is considered as a single point.

The author wishes to state that he benefited greatly from conversations with Dr. G. E. Schweigert during the earlier parts of the work.

Part I. Development of the theory

Let C be a cyclicly connected continuous curve which is not a simple closed curve.

Lemma 1. *Every free arc* of C is a subset of a maximal free arc of C.*

Proof. Let α be a free arc and a be an interior point of α. Let J be a simple closed curve containing a. Obviously $J \supset \alpha$. On each of the two arcs of J from a let x and y be the first point of the closed set $\overline{C-J}$. Since $C \neq J$, x and y exist. Further $x \neq y$, for otherwise this single point would be a cut point of C. Then xay is a free arc containing α and maximal.

We now define the following terms:

(a) A 2-*set* consists of any pair of points of C which separate C and do not lie together on the same free arc of C.

(b) A 2-*point* of C is any point h belonging to a 2-set of C.

(c) Two points a and b of C are *biconjugate* if no 2-set separates C between a and b.

(d) By a *secondary element of* C we shall mean any subset E of C satisfying some one of the following conditions:

I. E consists of a non 2-point a of C together with all points b of C biconjugate to a, and there is at least one such point b. In this case E is called a *true secondary* element. A true secondary element is said to be *nondegenerate* provided that it contains at least one nondegenerate component.

II. E is a non 2-point a such that no point b of C is biconjugate to a. In this case a is called a 2-*end point*.

III. E is a maximal free arc of C containing at least one 2-point in its interior.

IV. E is a 2-set.

* An arc axb of C is called a free arc of C provided the interior of axb, which we denote by $\langle axb \rangle$, is an open subset of C. We shall denote the sets $axb-a$ and $axb-b$ by $\langle axb$ and $axb\rangle$, respectively. An arc axb is said to span a point set M provided $axb \cdot M = a+b$. This lemma is an unpublished result of G. C. Watson. The proof here given was suggested by the referee.

The following theorems are easily established:

THEOREM I. (a) *The space is the sum of its secondary elements, in fact every point p of C is contained in a secondary element of type* I, II, *or* IV.

(b) *Every secondary element is a closed set.*

(c) *Every two points of a true secondary element are biconjugate.*

(d) *Every true secondary element is independent of the non 2-point defining it, that is, it may equally well be defined by any non 2-point which it contains.*

(e) *Every non 2-point belongs to at most one true secondary element.*

(f) *The product of two true secondary elements is vacuous, a 2-point, or a 2-set.*

(g) *If A be a free arc of C with end points a and b containing at least one 2-point r in its interior, then every 2-set K of C which contains r separates C between a and b.*

(h) *If A be a free arc of C, then a necessary and sufficient condition that A contain a pair of biconjugate points is that every interior point of A be a non 2-point.*

(i) *If E be a true secondary element, then every interior point of every free arc of E is a non 2-point.*

(j) *If E be a secondary element of type* III, *then every interior point of E is a 2-point and E contains no biconjugate pair of points.*

(k) *If E be a true secondary element containing b and c, then E contains every free arc joining b and c in C.*

THEOREM II. *Let E be a true secondary element and K any component of* $C-E$. *Then*

(a) $F(K)=\overline{K}-K$ *consists of exactly two points, r and s;*

(b) *both r and s are local separating points of C of order greater than* 2;

(c) *either r or s is a 2-point.*

Proof. (a) Since C is cyclic, $F(K)$ contains at least two points x, y. Let z be any third point of $F(K)$. Let U and W be regions containing x and z, respectively, and having disjoint closures neither of which contains the point y. Let x', z' be arbitrary points in KU, KW. There exists a simple arc $x'z'$ in K; hence $x'z'$ does not contain y. Let V be a region in C, containing y, such that $\overline{V}$ is disjoint with each of the three sets $\overline{U}$, $\overline{W}$, and $x'y'$. Let $x'x$, $z'z$ be simple arcs in U, W, respectively, and x'', z'' the first points of E on these arcs (Theorem I(b)). It follows easily that there exists a simple arc $x''z''$, having its interior in K, which does not intersect $\overline{V}$. We may find a point c in K and three simple arcs cx'', cy'', cz'', disjoint except for c, having their end points x'', y'', z'' in E, but otherwise lying in K. Now let a be any non 2-point of E. Then c is not biconjugate to a, hence there exists a 2-set H separating C be-

tween a and c. Let L be the component of $C-H$ which contains c. Since H consists of two points, L must contain one of the three arcs cx'', cy'', cz'', say cx''. Thus x'' is not biconjugate to a, which contradicts Theorem I(c). This proves (a).

(b) Evidently r and s are both local separating points of C. Neither of these points is of order one in C, since C is cyclic. Hence (b) will be proved if we show that r is not of order 2, since the same proof will hold for s. We first observe that E contains a non 2-point distinct from both r and s. If (r, s) is a 2-set, this is immediate. Otherwise there exists a free arc rs in C. By Theorem I(k), rs is in E, hence by Theorem I(i) every interior point of rs is a non 2-point. Therefore, using Theorem I(d), we may assume that the non 2-point a used to define E is neither r nor s. Let z be any point of K and define

$$4d = \min\,(\rho(a, r), \rho(z, r), \rho(r, s)) \geqq 0.$$

Assuming that r is of order 2 in C, let $U(r)$ be a neighborhood of r in C of diameter less than d, having two points y and w as its boundary. From the definitions of K and d we may assume without loss of generality that w lies in K and y in $C-\overline{K}$. Thus (w, y) separates C between a and r. Since r is in E, it follows by Theorem I(c) that there must exist a free arc wy in C. We have at once that either r or s is interior to the free arc wy. Call the one of these points satisfying this condition t. Now w is not biconjugate to a, hence there exists a 2-set (h, k) separating C between w and a. Since t is in E, (h, k) does not separate C between a and t. Thus either h or k is t or (h, k) separates C between t and w. Hence in either case one of the two points h, k, say h, must be interior to the free arc wy. Thus k does not lie on wy, and we see at once that if p is any point interior to wy then (p, k) is a 2-set separating C between a and w. Taking p interior to the subarc ty of wy, we have at once that (p, k) separates C between a and t, which is impossible since t lies in E. The proof of (b) is thus complete.

(c) We postpone the proof of this part of Theorem II until after the proof of Theorem VI(a) in order to establish some results which we shall need in its demonstration. To this end let R be the collection of all 2-points of C; L the collection of all non im kleinem cut points* of C; and W the set of all ramification points of C.

LEMMA 2. *Every nondegenerate subcontinuum G of R contains a free arc of C.*

Proof. Since G consists entirely of local separating points of C it follows†

* For definitions of the new terms used see G. T. Whyburn, Mathematische Annalen, vol. 102 (1930), pp. 313–336. We shall use Theorem 10, p. 320, of this paper and refer to it as Theorem B: Let T be any arc of C. Then if W is dense on T, LT has at least the power of the continuum.

† See G. T. Whyburn, Bulletin of the American Mathematical Society, vol. 39 (1933), p. 97.

that G is a locally connected continuum. Let T be any simple arc in G. It follows* that T contains at most a countable number of points of order different from 2.

Since T consists entirely of local separating points of C, it follows by Theorem C that every point of T is an im kleinem cut point. Thus $LT=0$. Therefore, W is not dense on T, by virtue of Theorem B. Hence T contains a subarc T' which is a free arc of C. Since T' is contained in G, the proof is complete.

THEOREM III. (a) *Every continuum of convergence K of C is contained in some true secondary element E of C. Moreover, every such K contains a non 2-point of C.*

(b) *If (K_i) be an infinite sequence of disjoint subcontinua of C having a nondegenerate limit inferior L, then there exists a true secondary element G of C which contains L, and is such that $L=\liminf K_iG$.*

Proof. (a) Since K is a continuum of convergence of C, no two points of K can be separated in C by a finite set. Thus every two points of K are biconjugate and, in particular, K contains no free arc. Hence the proof is complete if we show that K contains at least one non 2-point of C. This fact follows at once from Lemma 2.

(b)† It follows from the disjointness of the K_i that L is a continuum of convergence of C, hence by (a) there exists a true secondary element G of C which contains L. Let L' be the limit inferior of (K_iG). We have at once that L' is contained in L and need only show the reverse.

To this end let q be any point of L, and (q_i) any sequence of points converging to q such that q_i lies in K_i for each i. If q is not in L', we may assume without loss of generality that q_i lies in K_i-G for every i. Let S_i be the component of $C-G$ containing q_i. By definition q lies in $F(\sum S_i)$, and‡

$$F(\sum S_i) \subset \overline{\sum F(S_i)}.$$

Assume for the moment that there exists an S_i, call it S, which contains infinitely many of the sets K_i. It follows at once that every point of L is a boundary point of S, contrary to Theorem II(a).

We shall show in Theorem IV that for any positive number d there can

* See G. T. Whyburn, Monatshefte für Mathematik und Physik, vol. 36, pp. 305–314. We shall use Theorem 9 of this paper and call it Theorem A: All save possibly a countable number of the local separating points of a continuum M are points of order 2 of M. We shall also refer to Theorem 8 of this paper as Theorem C: The properties of being an im kleinem cut point and of being a local separating point are equivalent for all points of a continuum M at which M is locally connected.

† See Kuratowski and Whyburn, loc. cit. The present proof is a modification of the proof of (6, 7) of this paper.

‡ See Hausdorff, *Mengenlehre*, Berlin, 1927, p. 155.

exist at most a finite number of S_i having diameter greater than d. From this result and the previous paragraph it follows easily that we may pick a subsequence of the S_i, which we may assume is the whole sequence, consisting of distinct elements, and then a subsequence of the K_i, which we take as the whole sequence, such that each K_i contains a point r_i not in S_i and a point q_i in S_i. It follows at once that for every i the set K_i contains a point p_i of $F(S_i)$. Hence p_i converges to q since q_i converges to q. But p_i lies in $G \cdot K_i$, for each i; thus q belongs to L', and the proof will be complete when we establish Theorem IV.

THEOREM IV. *If E be a true secondary element, then there exists for every positive number d at most a finite number of distinct components of $C-E$ having diameter greater than d.*

Proof. Otherwise there exists an infinite sequence (S_i) of distinct components of $C-E$ each of diameter greater than d. We lose no generality in assuming that (S_i) is a convergent sequence to the limit set L. From the fact that E is closed (Theorem I(b)) we deduce immediately that L lies in E. Since L is a connected set of diameter at least d, let a, b, c be any three distinct points of L. Choose U, V, W as regions in C having disjoint closures and containing a, b, c respectively. Let S_k be an S_i intersecting all three of the regions U, V; W. It then follows easily that S_k has at least three boundary points, contrary to Theorem II(a).

This completes the proof of Theorems III and IV.

THEOREM V. *If E be a true secondary element and N a component of $C-E$, then for every true secondary element F we have either $FN=0$ or $\overline{N} \supset F$.*

Proof. Let a and b be non 2-points defining E and F respectively. Then there exists a 2-set (h, k) which separates C between these two points. It follows easily that $C-(h, k)$ has exactly n components A_i, $(i=1, 2, \cdots, n)$, for some finite n. We lose no generality in assuming that b lies in A_1 and a in A_2. It follows at once that F is contained in $\overline{A}_1$, and E is contained in $\overline{A}_2$. But $C-E$ contains $C-\overline{A}_2=A_1+\sum_3^n A_i$; hence if N is the component of $C-E$ containing b we have that $\overline{N}$ contains $\overline{A}_1$ which contains F. Thus $\overline{N} \supset F$. Evidently, $FM=0$ for every other component M of $C-E$. This completes the proof of Theorem V.

LEMMA 3. *If E be a true secondary element, and K a component of $C-E$ with boundary points r and s, then $\overline{K}$ contains at most one true secondary element F of C containing $r+s$.*

Proof. Assume that there exist two such true secondary elements F and G of C. By Theorem I(f), (r, s) is a 2-set; hence both r and s are 2-points. Let a

and b be non 2-points defining F and G, respectively, and let axb be a simple arc in K. Let w be the last point of axb in F and z the first point of axb following w which is in G. By Theorem I(f), w and z are distinct points; hence the arc wz is nondegenerate. If M be the component of $C-G$ containing a, then $\overline{M}$ contains F by Theorem V. It follows that M has the three boundary points r, z, s, contrary to Theorem II(a). The proof is thus complete.

THEOREM VI. (a) *For every positive number d there exists at most a finite number of true secondary elements E having components of diameter greater than d.*

(b) *C contains at most a countable number of distinct true secondary elements.*

Proof. Contrary to the theorem, let (E_i) be an infinite sequence of distinct true secondary elements having components of diameter greater than d. The proof now splits into two cases:

CASE (i). *For each i the set E_i intersects at most a finite number of the sets E_j.*

In this case there is an infinite subsequence of the E_i no two of which have a point in common, and we may assume this is the entire set E_i. Let e_i be a component of E_i of diameter greater than d. We may assume that the sequence (e_i) converges to a limit set e of diameter at least d. From Theorem III it follows that there exists a true secondary element F of C containing e. We may assume F is distinct from all the E_i. Using Theorems IV and V, we may assume that there exists a component N of $C-F$ such that $\overline{N}$ contains E_i for every i. Using Theorem II(a) we may assume that all the E_i are disjoint with F. Hence all the (e_i) are contained in N, a single component of $C-F$ which gives an easy contradiction to Theorem II(a) since C is a continuous curve. Thus the theorem is proven for Case (i).

CASE (ii). *There exists an E_i intersecting infinitely many E_j.*

We may assume in this case that E_1 intersects all the remaining E_i, and, by Theorems IV and V, that there exists a component N of $C-E_1$ such that $\overline{N}$ contains all the remaining E_i. Using Theorem II(a), let r and s be the boundary points of N. By Lemma 3 we may assume that every E_i contains r and no E_i, other than E_1, contains s. Continuing, this process of taking subsequences indefinitely we may assume without loss of generality that for i not j we have $E_iE_j=r$. Let q lie in $E_1-(r, s)$ and define $5m=\rho(q, \overline{N})$, $5n=d$ and $k=\min(m, n)$. If we define $d'=3k$, we have at once that d' is less than d.

Let e_i be a component of E_i having diameter greater than d'. It follows easily that there exists a point p_i in e_i, such that $\rho(p_i, r)$ is greater than k. Let V_i and U, respectively, be the spherical k-neighborhoods of p_i and q; and

define G_i as the sum of these neighborhoods. We shall assume, henceforth, that i is not one. We note in passing that $\overline{U}$ and E_i are disjoint for all such i. We now know that e_i is a continuum and G_i is an open set intersecting both e_i and $C-e_i$. It follows, therefore, by a well known theorem* that the component K_i of $\overline{G}_i e_i$ which contains p_i contains at least one point of $F(G_i)$, and hence at least one point of $F(V_i)$. Thus K_i is a continuum of diameter at least k which does not contain r. The K_i, being subsets of the corresponding E_i, are disjoint by pairs. The proof given for (i) now holds exactly if we substitute the sets K_i for the sets e_i used in that proof, and k for d. This leads to the same contradiction and completes the proof of (a).

We postpone the proof of (b) until after the development of some special theory of degenerate true secondary elements. Meanwhile we give the following proof.

Proof of Theorem II(c). Assuming the theorem false we first prove

(i) *K contains a continuum of convergence of itself. Furthermore, every simple closed curve J containing r, s and a point of K intersects some continuum of convergence of K in at least two points. Moreover, the true secondary element F of C which contains this continuum of convergence is contained entirely in K.*

Proof. Let J be any such simple closed curve and let A be any simple arc of J joining r and s in $\overline{K}$. Since (r, s) is not a 2-set, but does separate C, it follows that there exists a free arc $B=rxs$ in C, and B is in E by Theorem I(k). Thus if we define $J'=A+B$, it follows that J' is a simple closed curve such that $J'K$ is contained in JK. Hence it suffices to prove the theorem using J' in place of J.

From our hypothesis and Theorem I(i) it follows that every point of B is a non 2-point, hence no point of K is biconjugate to any point of B. Let x be an interior point of A; then there exists a 2-set (h_1, k_1) separating r from x in C. It follows at once that both h_1 and K_1 are interior to A and that we may assume that h_1 lies between r and x and k_1 between x and s on this arc.

Now (r, k_1) does not separate C. If it did, since r is a non 2-point, we could find a free arc rk_1 in C. It follows easily that this gives a theta curve in C having r and k_1 as vertices. This is impossible since k_1 is not biconjugate to r.

Since (r, k_1) does not separate C, there exists a simple arc W joining x to a (an interior point of B) but containing neither r nor k_1. Since (h_1, k_1) separates C between x and s, and since no interior point of A is biconjugate to s, we may find a simple arc T_1 in K spanning J' and having its end points a_1, b_1 interior to the subarcs rh_1 and sk_1 of A, respectively, except that a_1 may possibly be h_1.

* See Hausdorff, loc. cit., p. 161.

Now a_1 is not biconjugate to r; hence there exists a 2-set (h_2, k_2) separating a_1 from r in C, and we may assume that h_2 is interior to ra_1 and k_2 is on the half open arc $sb_1\rangle$. Using the argument of the preceding paragraphs, we may find a simple arc T_2 in K spanning J' between a_2 and b_2. These points will lie respectively in the half open arc $\langle rh_2$ and the open arc $\langle sk_2\rangle$. It follows at once that (h_2, k_2) separates C between every point interior to T_1 and every interior point of T_2 whence an easy consequence is that T_1, T_2 are disjoint.

In the same manner we may find an infinite sequence (T_i) of disjoint simple arcs in K spanning J' each between a point of rh_1 and a point of sk_1; hence all of the T_i are of diameter as least as great as the distance between these disjoint closed sets. Thus we may assume that the T_i converge to a non-degenerate limit set T. Hence by Theorems III(a), I(f), V, the fact that both r and s are non 2-points, and the following assertion (which follows easily from Theorems II(a) and III(b)), there must exist a true secondary element F of C which contains T and is contained in K.

ASSERTION. *If E be a true secondary element, then there exists no convergent infinite sequence of disjoint subcontinua of $C-E$ each member of which is of diameter greater than some preassigned d such that the limit of this sequence is in E.*

From the compactness of rh_1 and sk_1 we have at once that T must intersect the interiors of both of these arcs. This completes the proof of (i).

We continue to use the simple closed curve J' defined in proving (i). We see by that Theorem that K contains a continuum of convergence t_1 of itself which is contained in a true secondary element E_1 of C. Moreover, by (i), E_1 is contained in K and has at least two points in J'. It follows at once that these points must both be interior to A. Let m_1, n_1 be respectively the first and last points of the closed set E_1 in A. Define

$$d = \min (\text{diam } t_1, \rho(rm_1, sn_1)).$$

We shall now prove the following assertion:

(ii) (r, m_1) *does not separate* C.

Proof. Otherwise, since r is a non 2-point, there exists a free arc rm_1 in C. By Theorem II(b), s is not interior to this free arc; hence $\langle rm_1$ is contained in K. Let z be any point of $\langle rm_1\rangle$. Then (s, z) separates C between r and m_1. Otherwise there exists a simple arc rwm_1 in $C-(s, z)$, and it follows easily that $\langle rwm_1$ lies in K. Since rwm_1 does not contain z, it must be disjoint with the interior of the free arc rm_1. It is easily established from these facts that K contains a point w which can be joined to r by three independent arcs of C. This is impossible since no point of K is biconjugate to r.

Therefore, (s, z) separates C. Thus, since s is a non 2-point, there must exist a free arc sxz in C. This free arc cannot contain r by Theorem II(b), hence it must contain m_1. But the half open arc $rym_1\rangle$ is contained in a single component N of $C-E_1$, by definition of m_1. This gives a contradiction to Theorem II(b), and thus establishes (ii).

It is immediate that there exists a component N of $C-E_1$ such that $F(N)=(m_1, n_1)$; thus (m_1, n_1) separates C. Therefore, by (ii), there exists a simple arc xy in K spanning J', where x is in $\langle rm_1\rangle$ and y in sn_1. Thus x is biconjugate to y, whence y is not s.

Using the notation of (i) we find a 2-set (h_1, k_1) separating r from x in C, where h_1 is in $\langle rx\rangle$ and k_1 in $\langle sn_1\rangle$. It follows at once that (h_1, k_1) separates C between r and n_1. Continuing as in (i) we obtain a continuum of convergence t_2 of K intersecting both sk_1 and $\langle rh_1\rangle$. We let E_2 be the true secondary element of C containing t_2, and note that E_2 is contained in K and is of diameter at least d. Also, E_1 and E_2 are distinct. Repeating this process indefinitely gives an infinite sequence (t_i) of continua of convergence of K, each lying in a different true secondary element of C and each of diameter at least d. This contradicts Theorem VI(a) and completes the proof of Theorem II(c).

THEOREM VII. *If E be a degenerate true secondary element, then every two points p and q of E are vertices of a theta curve in C.*

Proof. By a theorem of N. E. Rutt* our theorem is equivalent to showing that no pair of points (h, k) separates p and q in C. Assuming it false, let J be any simple closed curve in C containing p and q; then J contains h and k. By Theorem I(c) there exists a free arc hxk in C.

CASE (i). *p is a non 2-point.*

If hxk contains p, then by Theorem I(j) the free arc hxk is contained in E, which is impossible since E is degenerate. If hxk contains q, then, since q is biconjugate to p, every interior point of hxk is biconjugate to p, and we are led to the same contradiction through Theorem I(d).

Thus hxk spans J, whence h is biconjugate to k, and by Theorem I(h), every interior point c of hxk is a non 2-point. Thus hxk lies in a nondegenerate true secondary element F of C. By Theorem I(f), c is not in E; hence c is not biconjugate to p. This yields an easy contradiction to the fact that p and q are biconjugate points, since any 2-set which separates c from p in C will also separate p from q. Thus Case (i) is established.

* See N. E. Rutt, American Journal of Mathematics, vol. 51 (1929), pp. 217–246, where the theorem is established for the plane. (Note in particular Theorem 4, p. 244, of this paper.) For the proof of the theorem for more general spaces see G. Noebeling, Fundamenta Mathematicae, vol. 18 (1932), p. 23–38.

CASE (ii). *Both p and q are 2-points.*

There exists a point z such that (p, z) is a 2-set. By Theorem I(c) there exists a component K of $C-(p, q)$ such that $\overline{K}$ contains E; hence K contains a non 2-point a of E. Let b be any point of $C-\overline{K}$, and J any simple closed curve in C containing a and b. Thus J contains both p and q. The half open arc $\langle pb$ of J is contained in $C-\overline{K}$ and hence in $C-E$. Thus p is a boundary point of some component of $C-E$. Hence by Theorem II(b), p, and similarly q, is a local separating point of C of order greater than 2. Thus hxk spans J; hence h is biconjugate to k, and, by Theorem I(h), every interior point of hxk is a non 2-point. Let c be any interior point of hxk and F the true secondary element of C containing c. Since p is biconjugate to q, it follows easily that both p and q are biconjugate to c, hence, Theorem I(f), $EF=(p, q)$ and (p, q) is a 2-set. F is also nondegenerate, when considered as a true secondary element. If b be a non 2-point defining E, it follows at once that b is not on the free arc hxk. Thus (h, k) separates C between b and either p or q, say p. Let J' be a simple closed curve in C containing p and b; then J' contains h and k and is spanned by the simple arc hxk. It follows that b is biconjugate to c, since p is biconjugate to b. Hence EF contains the three distinct points p, q, b contradicting Theorem I(f) and proving the theorem.

In the first paragraph of the proof of Theorem VII we established the following theorem.

THEOREM VIII. *Every 2-point p of C which is contained in a true secondary element E of C is a local separating point of C of order greater than 2 and also a boundary point of some component of $C-E$.*

THEOREM IX. *If E be a true secondary element containing no free arc, and K is any component of $C-E$, then $F(K)$ is a 2-set.*

The proof is immediate from Theorems I(k), II(a), and II(b).

THEOREM X. *If E be a degenerate true secondary element, then E contains infinitely many 2-sets and $C-E$ contains an infinite sequence of components whose closures are disjoint by pairs.*

Proof. Let a be any non 2-point of E and b an arbitrary point of $E-a$. By Theorem VII there exists a theta curve $H=\sum_1^3 ax_ib$ in C having a and b as vertices. Using Theorem I(k) and the fact that a is a non 2-point, we have at once that $C-(a, b)$ is connected. It follows that there is an arc h_1k_1 spanning H between ax_1b and one of its other two arcs. We lose no generality if we assume that h_1 lies in $\langle ax_1b\rangle$ and k_1 in $\langle ax_2b\rangle$. It is immediate that both h_1 and k_1 are in E since both of these points are biconjugate to a.

We have thus shown that for any point b in $E-a$ and any theta curve H in C having a and b as vertices, there exists a point q (h_1 above) interior to ax_1b such that q and a are vertices of a theta curve in C.

Using h_1 as b we may find an h_2 in $\langle ax_1h_1\rangle$ such that h_2 is in E and h_2 and a are vertices of a theta curve in C. Continuing this process indefinitely we may construct an infinite sequence of points (h_i) such that for every i we have

(a) h_i is in E;

(b) a precedes h_{i+1} precedes h_i precedes b on ax_1b;

(c) h_i and a are vertices of a theta curve in C.

We may assume that the sequence (h_i) converges to a point h. We have that a lies on ax_1b and precedes all the h_i on this arc.

Since E is degenerate there exists for each n a point p_n in the subarc $\langle h_{2n}xh_{2n-1}\rangle$ of ax_1b which is not in E. Hence, (Theorem I(b)), let w_n, z_n be respectively the last point of E preceding and the first point of E following p_n on ax_1b. We have at once that both of these points are contained in the closed arc $h_{2n}x_1h_{2n-1}$; hence if K is the component of $C-E$ containing p_n it follows from Theorem II(a) that $F(K_n)=(w_n, z_n)$.

We have at once that the infinite sequence of components (K_{2i-1}) of $C-E$ are mutually separated by pairs. The remainder of the theorem follows at once from Theorem IX.

Proof of Theorem VI(b). By virtue of Theorem VI(a) it suffices to prove that C contains but a countable number of degenerate true secondary elements. Assume the theorem false and let (E) be any uncountable collection of degenerate true secondary elements of C. Let

$$p_1, p_2, p_3, \cdots \tag{1}$$

be the set of all local separating points of C contained in at least one of the sets (E). This set is countable by Theorem A. Using Theorems VIII and X, we see that every set E contains infinitely many distinct points of (1). It follows that some point of (1), say p_1, must be contained in an uncountable subcollection of the (E); hence we lose no generality in assuming that it is contained in all the E. Let E_1 be any one of these sets. From Theorems IV and V we lose no generality in assuming that there exists a component K_1 of $C-E_1$ such that $\overline{K}_1$ contains all the sets (E) except E_1. Using Theorem II(a) and Lemma 3, we may let $F(K_1)=(p_1, s_1)$ and assume that s_1 lies in none of the sets E other than E_1. From Theorems VIII and X every E contains a 2-point q distinct from p_1. From this it follows easily that we may assume p_2 lies in uncountably many E; hence uncountably many E's contain both p_1 and p_2. This gives an easy contradiction to Lemma 3 using Theorem IV. The proof of Theorem VI(b) is thus completed.

THEOREM XI. *Every isolated point q of a true secondary element E is a 2-point.*

Proof. Let a be any non 2-point of E and akq a simple arc in C. Let k be the last point of E preceding q on akq. It follows at once from Theorem II(a) that there exists a component K of $C-E$ such that $F(K)=(k, q)$. The theorem is then immediate by Theorem I(k).

Notation. For any set M let $L(M)$ denote the number of components of M, if $M \neq 0$, and $L(M)=1$ if $M=0$.

THEOREM XII.* *Let F be a closed subset of a compact continuum S such that each component of $S-F$ has a most n limit points in F. Then for any continuum N of S, we have*

$$L(NF) \leqq nL(S-F).$$

Proof. We have only to consider the case where $L(S-F)$ is finite and $N \not\subset F$, as the inequality is obvious in the other cases. Let P be the set of all limit points in F of components of $S-F$. Then the number of points in P is less than $nL(S-F)$. Let q be a point of $NF-P$. By a theorem of Hausdorff the component of N containing q and lying in the open set $U=S-\overline{S-F}$ has a limit point on $F(U)$. But this component belongs to $NF-P$ and $F(U) \subset P$. Thus each component of NF contains a point of P which proves the theorem.

COROLLARY. *For any true secondary element E of C and any continuum N in C we have*

$$L(NE) \leqq 2L(C-E).$$

THEOREM XIII. *If $Z(E)$ is finite, and K is a continuum of convergence of C which is contained in E, then K contains a continuum of convergence of E.*

Proof. Let $Z(E)=h$, and let d be so chosen that the diameter of K is $4hd$. Since K is a continuum of convergence of C, there exists an infinite sequence (K_i) of disjoint subcontinua of C converging to K. By Theorem III(b) the sets K_iE will also converge to K. We thus lose no generality if we assume that the diameter of K_iE is greater than $3hd$ for all i. Since K is a continuum of diameter $4hd$, there will exist $3h$ points p_i, $(i=1, 2, \cdots, 3h)$, in K such that $\min \rho(p_i, p_j)=d$, $(i \neq j)$. Construct the neighborhoods $V_i(p_i)$, $(i=1, 2, \cdots, 3h)$, in C having disjoint closures. For i sufficiently large K_iE will intersect all of these neighborhoods. By the corollary to Theorem XII, K_iE has at most $2h$ components for every i. It follows that for every sufficiently large i some component k_i of K_iE must intersect at least two of the neighborhoods $V_i(p_i)$. Using this fact and the information that the $V_i(p_i)$

* This theorem and its proof were suggested by the referee.

have disjoint closures, the required continuum of convergence of E is easily constructed.

Part II. Secondary extensibility and reducibility

Definition. *Let P be a property of point sets. Then P will be said to be secondarily extensible provided C has property P whenever every component of each true secondary element of C has property P. The property P will be said to be secondarily reducible provided the assertion C has property P implies that every component of each true secondary element E of C has property P.*

Theorem XIV. *Let C have the property that for each true secondary element E of C, $L(C-E)$ is finite. Let P be any one of the following properties*: (1) *being hereditarily locally connected*; (2) *being a regular curve*; (3) *being a rational curve. Then P is secondarily extensible and reducible.*

Proof. The proof that each of these properties is secondarily reducible follows at once from the fact that if C has property P then every subcontinuum of C necessarily has property P.

(1) *is secondarily extensible.*

Otherwise there exists a continuum of convergence K of C. Using Theorems III(a) and XIII we obtain at once a continuum of convergence of some component of a true secondary element E of C. This contradiction establishes our assertion.

(2) *and* (3) *are secondarily extensible.*

Assume that C is not a regular (rational) curve. Then, by a theorem of Menger,* C contains a subcontinuum N composed entirely of irregular (irrational) points of C. Using two theorems of G. T. Whyburn,† we see that the non-local separating points of C are dense on N. Let a be any non-local separating point of C contained in N. Then a is a non 2-point of C. Since a is of order greater than 2, it is not a 2-end point of C and hence defines a true secondary element E of C. Since $Z(E)$ is finite and a is not a local separating point of C, it follows that a is not a limit point of $C-E$. Hence by Theorems XI and XII we see that a lies in a nondegenerate component K of NE. Let M be the component of E containing K. From the Corollary to Theorem XII we have that a is not a limit point of $C-M$, hence the order of a in M is the same as the order of a in C. Therefore, a is an irregular (irrational) point of M. Thus M is not a regular (rational) curve. This contradiction establishes (2) and (3).

* See K. Menger, *Kurventheorie*, Leipzig, 1932, pp. 127, 133.

† See Bulletin of the American Mathematical Society, vol. 35 (1929), p. 102, Theorems 13 and C.

THEOREM XV. *The property of being a regular curve is not secondarily extensible.**

Proof. The proof will be by an example constructed in the Euclidean plane P. Let ABD be an equilateral triangle in P with base AD. Let E be the midpoint of AD and join E to the midpoints of AB and BD by line segments EF and EG respectively. There will then be two triangles (namely AFE and EGD) having their bases on AD. Repeat the above construction on each of these triangles, and continue the process indefinitely. Call the resulting configuration K Let K' be the reflection of K in AD with B' the point corresponding to B under the reflection. Through each point of AB' having order three construct a segment parallel to AD and extended until it meets the segment $B'D$. Denote by H the set K' modified by the addition of these intervals. Let C' consist of H, K and a free arc BB' formed by two segments and spanning the point set $H+K$. Let t be any free arc of $H+BB'$ and p any interior point of t. With p as center and sufficiently small radius we may draw a circle disjoint with $\overline{C'-t}$, but lying in the plane of C'. Let C be the set C' modified by the addition of such a circle to every free arc of $H+BB'$. Then the secondary elements of C may be classified as follows: (1) one disconnected nondegenerate true secondary element G consisting of K and all points of $H+BB'$ of order greater than two; (2) a countable number of connected nondegenerate true secondary elements, each consisting of a theta curve composed of a circle and a proper subinterval of some free arc of $H+BB'$; (3) a countable number of secondary elements of type III each consisting of the closure of a component of the complement in C of all secondary elements mentioned in (1) and (2).

It will be noted that although no secondary element of C contains an irregular point of itself, the set C is not even hereditarily locally connected.

PART III. EXAMPLES

THEOREM XVI. *There exist true secondary elements E having each of the following properties*:

(a) *E is connected and locally connected*;
(b) *E is connected but not locally connected*;
(c) *E is locally connected but not connected*;
(d) *E is countable*;
(e) *E is uncountable*;
(f) *E is the closure of a free arc*;
(g) *E is cyclic*;

* Cf. Theorem XIV.

(h) *E has exactly one end point of itself*;

(i) *E is neither connected nor locally connected.*

Proof. (a), (e), (g): Let C be a 2-cell; then $E=C$.

For the remaining examples we define a *lumped free arc* of C to be on open subset A of C which contains B, where A is an open arc, B is a 2-cell, and $A \cdot B$ is an arc (closed).

(c), (e): Let C consist of two disjoint 2-cells H and K joined by three disjoint lumped free arcs; then $E=H+K$.

(e), (i): The example shall be constructed in the Cartesian plane. Let C_i be a circle and its interior having the center $(2^{-i}, 0)$ and the radius $1/2^{i+2}$. It follows easily that all the C_i are disjoint by pairs. Let p_i, q_i be respectively the points of C_i having the maximum and minimum ordinate. For each i let A_i be the line segment joining p_i to p_{i+1}, and B_i the line segment joining q_i to q_{i+1}. Let T denote the closed interval of the x axis from (0, 0) to (1/2, 0), and let D consist of the sum of all the sets A_i, B_i, C_i, T. In D replace each free arc by a lumped free arc and denote the resulting configuration by C. Let E denote the subset of C consisting of the origin and all the sets C_i. Then E is a true secondary element of C fulfilling the conditions (e) and (i).

(e), (f): Let C consist of a right triangle plus the three segments parallel to the hypotenuse and joining the points dividing its legs into four equal parts. Then E will be the middle one of these segments.

(d), (i): Let D' consist of an equilateral triangle pqr, having p at the origin and the side pq along the positive x axis, plus an infinite sequence of disjoint free arcs spanning pqr between points on pq and pr and converging to the point p. Let D'' be the reflection of D' in the y axis, where q' is the reflection of q and r' is the reflection of r. Define D as the sum of D' and D'' plus two disjoint spanning arcs of this set, one between q' and q, the other between r' and r. Let C be the configuration D in which every free arc has been replaced by a lumped free arc. Then E will consist of the non 2-point p plus the end points of all the lumped free arcs of C.

(a), (e), (h): In the Euclidean plane P let D be the sum of two disjoint 2-cells A and B. Let a, b, c be three distinct points of A which are limit points of $P-A$ and d, e two distinct points of B which are limit points of $P-B$. Let D' consist of D together with a spanning arc ad of D and a spanning lumped free arc ce of D which is disjoint with ad. Let f be any interior point of ad and bf a lumped free arc spanning D'. Let C consist of the entire configuration. Then E will be the sum of A and the free arc af of C.

(b), (e), (g): Let A be a square and its interior with center at p. Let B be a square and its interior with side half that of A, having its center at p

and its sides parallel to those of A. Let $D=\overline{A-B}$. Then D divides the plane into exactly two regions, the bounded one of which we call R. Thus the boundary of R is a square having sides W, X, Y, Z in the counterclockwise order. We define (D_i) as an infinite sequence of disjoint 2-cells contained in $\overline{R}$, converging monotonically to W, and each intersecting $\overline{R}$ in exactly two simple arcs, one on X and one on Z. (The 2-cells are most easily pictured as strips.) Add i disjoint lumped free arcs in R spanning E between points of D_i and D_{i+1} and equally spaced in the region between these 2-cells. Then the resulting space C is a locally connected continuum. It is easily seen that the set E consisting of D and all the sets D_i is a true secondary element of C satisfying (b), (e), and (g).

THE UNIVERSITY OF VIRGINIA,
CHARLOTTESVILLE, VA.

A CORRECTION TO "THE ASYMPTOTIC FORMS OF THE HERMITE AND WEBER FUNCTIONS"*

BY

NATHAN SCHWID

My attention has been called to the necessity of the following correction: in the formula for θ on page 353 the symbol $O(\kappa^{-3/2})$ must be replaced by $O(z^3\kappa^{-3/2})$, and in formula (26) and thereafter to formulas (27), as well as in (36)(a), the symbol $O(\kappa^{-1/2})$ is to be replaced by $O(z^3\kappa^{-1/2})$. With this modification, the z region in which the formulas (27) and (36)(a) will be found useful in any specific consideration is clearly more restricted than my introductory statement to §5 implies: such a region would be the one in which $|z| = o(\kappa^{1/6})$.

* Received by the editors December 26, 1939. Cf. these Transactions, vol. 37 (1935), pp. 339–362.

THE TEXAS COLLEGE OF MINES AND METALLURGY,
EL PASO, TEXAS.

LINEAR OPERATIONS ON SUMMABLE FUNCTIONS*

BY

NELSON DUNFORD AND B. J. PETTIS†

In the last few years the work of Gelfand [17, 18],‡ Kantorovitch [23, 24, 25], Dunford [9, 10], Vulich [24, 25, 42] and others has shown that in developing a representation theory for various classes of linear operations§ among Banach spaces [1] effective use can be made of abstract functions and integrals, just as the general linear functionals over certain B-spaces were earlier discovered to be representable in terms of numerical functions and the integrals of numerical functions. This is especially true for operations defined to a general B-space X from a Lebesgue space, that is, from a space consisting of a class of Lebesgue-integrable numerical functions. To obtain representations for operations of this sort it was found that ready application could be made of various integrals of the Lebesgue type that have been defined for functions taking their values in X.

In the present paper we wish to communicate a representation theory for several types of operators mapping a space $L(S)$, consisting of the real functions that are Lebesgue-integrable over an abstract aggregate S with respect to a fixed class of subsets of S and a fixed measure function [29, 34], into an arbitrary B-space X. The representations will be given in terms of abstract integrals and kernel integrals. The general approach is not new, for it is based on the methods introduced by Gelfand [18] and Dunford [9] to obtain such theorems when S is a bounded real interval. However, in order to extend these known results to the case of an arbitrary S new devices are required since the earlier results were proved by Euclidean methods. In most instances we have been able to make the extension; this has been accomplished by generalizing the Radon-Nikodym theorem [29, 34] to set functions taking their values in an adjoint space and by substituting for differentiation processes the use of convex sets. The class of operators recently introduced by Kakutani [22] and Yosida [44] under the name *weakly completely continu-*

* Presented to the Society, December 30, 1938, under the title *On completely continuous operations in L*; received by the editors October 12, 1939. A summary appeared in the Proceedings of the National Academy of Sciences, vol. 25 (1939), pp. 544–550.

† Sterling Fellow.

‡ Numbers in brackets refer to the references at the end. Gelfand's paper [18] is the thesis which he presented in June, 1935.

§ Hereafter for linear operations we shall use the briefer terms *operations* or *operators* since these are the only operations that will come into consideration. An operation, or operator, is thus understood to be a distributive continuous mapping of one B-space into another.

BOSTON UNIVERSITY
COLLEGE OF LIBERAL ARTS
LIBRARY

*ous** is also considered. By means of these representation theorems new information is given concerning certain types of operators. This information in turn yields a uniform mean ergodic theorem for weakly c.c.† operations in $L(S)$ and an application to Markoff processes. In addition it provides results which may be of interest in the theory of integral equations. In terms of both abstract integrals and kernel integrals a fairly complete representation theory is given for operations mapping $L(S)$ into the Lebesgue classes $L^q(T)$, $1 \leqq q \leqq \infty$, where T is another aggregate; the types considered are the general, the separable,‡ the weakly c.c., and the c.c. operations. Those results dealing with arbitrary S and T will have as immediate corollaries the corresponding theorems for the sequence spaces l^p;§ the supplying of these corollaries will be left to the reader.

A more precise outline of the contents is perhaps better given by the following table and comments.

Contents

Chapter I. Abstract functions and integrals

		Page
Part 1.	Classes of measurable and integrable abstract functions	325
A.	Preliminary definitions	325
B.	General measurable and integrable functions	327
C.	Measurable functions and absolutely integrable functions	329
Part 2.	Some convex sets related to integrable functions	331
Part 3.	Kernel representations of abstract functions and integrals	335
A.	Measurable functions in terms of numerical kernels	335
B.	The representation of abstract integrals by kernel integrals	337

Chapter II. General operations to restricted spaces

Part 1.	The representation in adjoint spaces of set functions and linear operations by abstract integrals	338
A.	Integral representations of absolutely continuous and Lipschitzean set functions	338
B.	Abstract representations of operations to certain adjoint spaces	345
Part 2.	Representations of separable operations to $L^q(T)$, $1<q\leqq\infty$	347
A.	Operations with range in $L^q(T)$, $1<q<\infty$	347
B.	Operations with range in $L^\infty(T)$	348
Part 3.	The representations of operators taking $L(S)$ into $C^*(T)$ or $L(T)$	351
A.	Operations to $C^*(T)$	352
B.	Operations from $L(S)$ to $L(T)$	358
Part 4.	Subclasses of those operations having their range in $C^*(T)$ or $L(T)$	358

* An operation is weakly completely continuous if it takes bounded sets into weakly compact sets, a set being weakly compact if every infinite subset contains a subsequence converging weakly to some element of the space.

† Hereafter *completely continuous* will be abbreviated to *c.c.*

‡ An operator is *separable* if it maps bounded sets into separable sets. Replacing "separable sets" by "compact sets" furnishes the definition for a *completely continuous* operator, which is necessarily separable. We shall also consider operations that send weakly compact sets into compact sets.

§ Concerning operations on or to a sequence space see [4, 9, 10, 18, 21, 23, 24, 31, 38] and references therein.

A. Operations to $C^*(T)$ given by almost separably-valued kernels. 359
B. Operations U' to $L(T)$ of the form $U'(\phi)=\int_S K'(s, t)\phi(s)d\alpha$. 362
C. Operations to $AC(T)$ given by almost separably-valued kernels. 363
Part 5. The iterative integrals associated with certain kernels. 364

CHAPTER III. RESTRICTED OPERATIONS TO ARBITRARY SPACES

Part 1. Weakly completely continuous and completely continuous operations. 366
A. Sufficient conditions for weak complete continuity and complete continuity. 367
B. Criteria for weak complete continuity and complete continuity when S is Euclidean. . 369
C. Abstract representations of completely continuous operations to $L^q(T)$, $1\leqq q\leqq\infty$. . . . 370
D. Approximation theorems. 375
Part 2. Applications. 376
A. Criteria for weak compactness in $L(S)$. 376
B. Kernel representations of operations among summable functions. 378
C. C.c. operations from $L^p(S)$ to X. 383
D. A mean ergodic theorem. 384
E. An application to Markoff processes. 388

Chapter I consists of groundwork. In Part 1 there is a rapid review of Lebesgue integrals for B-space-valued functions, Part 2 is explained by its caption, and Part 3 is chiefly spent in noting that previous results [9] concerning the representation (by kernels and kernel integrals) of measurable and integrable functions in certain function spaces are extensible from Euclidean S to arbitrary S. Most of Chapter II is devoted to establishing for an arbitrary S results known to hold in the Euclidean case; there are occasional refinements. It is believed however that Theorem 2.1.0, the first section of Part 3, and most of Part 4 are new. Chapter III is probably the most important. It includes the rather interesting result that when S is Euclidean and U is weakly c.c. in $L(S)$ the operation U^2 is c.c. This of course implies that U has a finite-dimensional set of fixed points. In Chapter III, Parts 2 A, C consist of sharpenings and extensions of well known theorems, and Part B contains a summary of our results (except for those in Part 3 of Chapter II) concerning kernel representations of operations on $L(S)$ to $L^q(T)$, $1\leqq q\leqq\infty$. These results enable us to state that several sets of conditions, which are less stringent than those heretofore considered, are sufficient that a kernel operation be c.c. from $L(S)$ to $L^q(T)$.

Chapter III is independent of Chapter II.

CHAPTER I. ABSTRACT FUNCTIONS AND INTEGRALS

PART 1. CLASSES OF MEASURABLE AND INTEGRABLE ABSTRACT FUNCTIONS

A. **Preliminary definitions.** We consider an arbitrary aggregate $S=[s]$, a fixed Borel field $\mathcal{E}=[E]$ of subsets of S with* $S\in\mathcal{E}$, and a fixed non-nega-

* To use Saks' term, E is an *additive family*.

tive completely additive measure function $\alpha(E)$ defined over $\mathcal{E}$. If we let $\mathcal{E}_B$ denote those $E \in \mathcal{E}$ for which $\alpha(E) < \infty$, a triad $[S, \mathcal{E}, \alpha]$ of this sort will be called a *system* provided that S is *decomposable*, that is, S has a *decomposition* consisting of a denumerable number $\{S_i\}$ of disjoint elements of $\mathcal{E}_B$ such that $S = \sum_i S_i$. Those elements of $\mathcal{E}$ for which $\alpha(E) = 0$ are the *null sets* and are denoted by $\mathcal{E}_0$. It is supposed that every subset of a null set is a null set. When the system $[S, \mathcal{E}, \alpha]$ and the exponent q are fixed, $L^q(S)$ is† the B-space composed of those numerical functions $\phi(s)$ which are defined over S, measurable with respect to $\mathcal{E}$, and have $\|\phi\| = (\int_S |\phi(s)|^q d\alpha)^{1/q} < \infty$ for $q < \infty$ and $\|\phi\| = \text{ess. sup. } |\phi(s)| < \infty$ for $q = \infty$. When $q < \infty$ we note that separability for $L^q(S)$ is equivalent to $\mathcal{E}_B$'s being separable under the metric dist $(E, E') = \alpha(E - E') + \alpha(E' - E)$. Moreover if $q < \infty$ then $f(\phi)$ is a linear functional over $L^q(S)$ if and only if an element ϕ' of $L^{q'}(S)$ exists such that $f(\phi) = \int_S \phi(s)\phi'(s) d\alpha$; the norm $\|f\|$ of f equals $\|\phi'\|$ and q' is the exponent conjugate to q.‡ Thus for $1 < q < \infty$ $L^q(S)$ is reflexive§ and hence [16] has a weakly compact unit sphere.

Given a function $x(s)$ defined from the points of S to a B-space X, the set of values assumed on a subset E of S will be denoted by $x(E)$. The function $x(s)$ is *separably-valued* if $x(S)$ is separable in X, *weakly compact-valued* if $x(S)$ is weakly compact, and *compact-valued* when $x(S)$ is compact. Should $x(S - E_0)$ be separable for some null set E_0 we say that $x(s)$ is *almost* (or *essentially*) *separably-valued*; the corresponding definitions for $x(s)$'s being *almost weakly compact-valued* and *almost compact-valued* are evident. More particularly, $x(s)$ is *finitely-valued* if it is constant on each of a finite number of measurable sets E_i with $\sum_i E_i = S$, and it is a *simple function* if it is finitely-valued and if $S - E \in \mathcal{E}_B$ when E is the set over which $x(s)$ vanishes. Finally, if Y is a subset of X, then $x(s)$ is *essentially defined to* Y if $Y \supset x(S - E_0)$ for at least one null set E_0.

In a given B-space X the zero element will be denoted by θ or θ_X. When Y is a subset of X the *span* of Y is the smallest closed linear manifold (c.l.m.) containing Y. Given Y in X and a set Γ in the adjoint X^* of X, Γ is said to be a *determining manifold*‖ *for* Y if a finite constant C exists such that, for each $y \in Y$, $\|y\| = \sup_\gamma |\gamma(y)|$, $\gamma \in \Gamma$, $\|\gamma\| \leqq C$. Γ is not required to be linear.

† We shall write $L(S)$ for $L^1(S)$. The elements of $L(S)$ will be called *summable functions*.

‡ When $\alpha(S) < \infty$ a proof of this statement, which is a classical result [1] for Euclidean S, can be found in [10, Theorem 46]; the induction to a general decomposable S is easily made. For the case $q = 1$ Nikodym was apparently the first to extend the classical theorem to functions of an abstract variable [30].

§ Even more, $L^q(S)$ is uniformly convex as shown by Clarkson's proof for Euclidean S, *Uniformly convex spaces*, these Transactions, vol. 40 (1936), pp. 396–414.

‖ When $Y = X$ the set Γ is simply a *determining manifold*.

Finally when a real function $\phi(s)$, or a B-space-valued function $x(s)$ or x_s, is being considered as an element of a class of functions, we shall usually write $\phi(.)$, $x(.)$, or $x.$, following the notation of E. H. Moore.

B. **General measurable and integrable functions.** Let $[S, \mathcal{E}, \alpha]$ be a fixed system and consider an abstract function $x(s)$ defined a.e.† in S and having its values in X. If Γ in some set in X^* and $f(x(s))$ is measurable for each $f \in \Gamma$, we say that $x(s)$ is Γ-*measurable*; for the case $\Gamma = X^*$ a Γ-measurable function is *weakly measurable*. To each p lying between 1 and ∞ inclusive there can be associated the subclass $\mathfrak{L}^p(S)[X, \Gamma]$ of Γ-measurable functions defined by the following condition: $x(.)$ is in $\mathfrak{L}^p(S)[X, \Gamma]$ if and only if $f(x(.)) \in L^p(S)$ for each $f \in \Gamma$. Denoting by p' the exponent conjugate to p, we recall the following fundamental theorem [17; 10, Theorem 49; 18].

THEOREM 1.1.1. *If Γ is a c.l.m. and $x(.) \in \mathfrak{L}^p(S)[X, \Gamma]$, a finite constant C exists such that*

$$\left| \int_S f(x(s))\phi'(s)d\alpha \right| \leqq C \cdot \|f\| \cdot \|\phi'\|, \qquad f \in \Gamma, \phi' \in L^{p'}(S),$$

C being independent of f and ϕ'. Hence

$$y_{\phi'}(f) \equiv \int_S f(x(s))\phi'(s)d\alpha$$

defines an operation $U(\phi') = y_{\phi'}$ from $L^{p'}(S)$ to Γ^*.

An element $x(.)$ of $\mathfrak{L}^p(S)[X, \Gamma]$ is said to be in the class $\mathfrak{L}_0^p(S)[X, \Gamma]$ if for each $E \in \mathcal{E}_B$ there is a point x_E in X such that

$$f(x_E) = \int_E f(x(s))d\alpha, \qquad f \in \Gamma; \tag{1}$$

when $p = 1$ it is supposed that such an x_E exists for each $E \in \mathcal{E}$. An element $x(.)$ of $\mathfrak{L}_0^1(S)[X, \Gamma]$ is a Γ-*integrable* function, and any point x_E satisfying (1) is a Γ-*integral of* $x(.)$ *over* E. If Γ is a determining manifold for X and $x(.)$ is Γ-integrable, then $x(.)$ has for each $E \in \mathcal{E}$ a unique Γ-integral over E called *the* Γ-*integral of* $x(.)$ *over* E. When $\Gamma = X^*$ this unique point is referred to as *the integral of* $x(.)$ *over* E and $x(.)$ is said to be *integrable*. Whenever in connection with a given abstract function $x(.)$ the symbol $\int_E x(s)d\alpha$ occurs devoid of qualification, the implication is that $x(.)$ is integrable and that $\int_E x(s)d\alpha$ stands for the integral of $x(s)$ over E.

The following is a slight extension of Theorem 59 of [10].

† The phrase *almost everywhere*, with its usual meaning of "except possibly on a set of measure zero," will be abbreviated throughout to *a.e.*

THEOREM 1.1.2. *When Γ is a determining c.l.m., $x(.)$ is in $\mathfrak{L}_0^p(S)[X, \Gamma]$ if and only if $x(.)\phi'(.)$ is Γ-integrable for each $\phi' \in L^{p'}(S)$.*

According to Theorem 1.1.1 if $x(.) \in \mathfrak{L}_0^p(S)[X, \Gamma]$ there exists a constant C independent of ϕ' and f and such that

$$\left| \int_E f(x(s))\phi'(s)d\alpha \right| \leqq C \cdot \|\phi'\| \cdot \|f\|, \qquad f \in \Gamma, \ \phi' \in L^{p'}(S). \tag{2}$$

At the same time it follows from the definition of $\mathfrak{L}_0^p(S)[X, \Gamma]$ that $x(.)\phi'(.)$ is Γ-integrable for each $\phi' \in L^{p'}(S)$ which is finitely-valued. Moreover, if for a ϕ' of this sort the Γ-integral of $x(s)\phi'(s)$ over E is $x_{E,\phi'}$, it is clear from (2) that

$$|f(x_{E,\phi'})| \leqq C \cdot \|\phi'\| \cdot \|f\|, \qquad f \in \Gamma,$$

and hence, since Γ is a determining manifold (for X),

$$\|x_{E,\phi'}\| \leqq C'\|\phi'\| \tag{3}$$

where C' is finite and independent of ϕ'. In addition $x_{E,\phi'}$ is additive over the finitely-valued elements of $L^{p'}(S)$ due to Γ's being determining. Now let $\{\phi_n'\}$ in $L^{p'}(S)$ be finitely-valued elements converging to an arbitrary $\phi' \in L^{p'}(S)$. It is clear from (3) that $\lim_{m,n\to\infty} \|x_{E,\phi_m'} - x_{E,\phi_n'}\| = \lim_{m,n} \|x_{E,\phi_m'-\phi_n'}\| = 0$. Let $x_{E,\phi'} = \lim_n x_{E,\phi_n'}$. Then

$$f(x_{E,\phi'}) = \lim_n f(x_{E,\phi_n'}) = \lim_n \int_E f(x(s))\phi_n'(s)d\alpha = \int_E f(x(s))\phi'(s)d\alpha, \qquad f \in \Gamma,$$

so that (1) is satisfied. Since $E \in \mathcal{E}$ was arbitrary, $x(.)\phi'(.)$ is Γ-integrable.

If on the other hand $x(.)\phi'(.)$ is Γ-integrable for every $\phi' \in L^{p'}(S)$, then for each $f \in \Gamma$ the function $f(x(s))\phi'(s)$ is summable for every $\phi' \in L^{p'}(S)$. Hence $f(x(.)) \in L^p(S)$ for each $f \in \Gamma$ [10, Theorem 46]. To see that $x(.) \in \mathfrak{L}_0^p(S)[X, \Gamma]$ we have only to observe that $x(.)\phi'(.)$ is Γ-integrable whenever ϕ' is the characteristic function of an element $E \in \mathcal{E}_B$ (or $\mathcal{E}$ if $p=1$).

THEOREM 1.1.3. *If Γ is a determining c.l.m. and $x(.)$ is in $\mathfrak{L}_0^p(S)[X, \Gamma]$, then $U(\phi') = x_{\phi'}$, where $x_{\phi'}$ is the Γ-integral of $x(.)\phi'(.)$ over S, is an operation defined from $L^{p'}(S)$ to X. The mapping $V(f) = f(x(.))$ is an operation from Γ to $L^p(S)$ that coincides over Γ with the adjoint of U.*

From Theorem 1.1.2 and (2) the Γ-integral $x_{\phi'}$ exists and satisfies the inequality

$$|f(x_{\phi'})| = \left| \int_E f(x(s))\phi'(s)d\alpha \right| \leqq C \cdot \|\phi'\| \cdot \|f\|, \qquad f \in \Gamma, \ \phi' \in L^{p'}(S), \tag{4}$$

where C is independent of ϕ' and f. The mapping V is obviously defined from Γ to $L^p(S)$ and is additive; it is therefore linear since from (4)

$$\|V(f)\| = \|f(x(.))\| \leqq C \cdot \|f\|, \qquad f \in \Gamma,$$

where C is independent of f. From (4) it also follows, since Γ is determining, that

$$\|x_{\phi'}\| \leqq C' \cdot \|\phi'\|, \qquad \phi' \in L^{p'}(S),$$

where C' is independent of ϕ'. Thus the additive mapping U from $L^{p'}(S)$ to X is linear. Finally, from the equality $f(U(\phi')) = \int_S f(x(s))\phi'(s)d\alpha$, $f \in \Gamma$, $\phi' \in L^{p'}(S)$, it is seen that V coincides over Γ with the adjoint of U.

If X is the adjoint Y^* of another space Y, it is evident from Theorem 1.1.1 that any element of $\mathfrak{L}^p(S)[Y^*, Y]$ is also in $\mathfrak{L}_0^p(S)[Y^*, Y]$. The space Y being equivalent to a determining c.l.m. in $X^* = Y^{**}$, Theorem 1.1.3 yields

THEOREM 1.1.4. *If $x(.) \in \mathfrak{L}^p(S)[Y^*, Y]$, then $x(.) \in \mathfrak{L}_0^p(S)[Y^*, Y]$ and the two mappings*

$$V(y) = x(s)(y), \qquad y \in Y,$$

$$U(\phi') = \int_S x(s)\phi'(s)d\alpha, \qquad \phi' \in L^{p'}(S),$$

where $\int_S x(s)\phi'(s)d\alpha$ is the Y-integral of $x(.)\phi'(.)$, are operations defined from Y to $L^p(S)$ and from $L^{p'}(S)$ to Y^* respectively. The operation V coincides over Y with the adjoint of U.

When Y is separable and $p = \infty$, it is possible to be more precise.

THEOREM 1.1.5. *If Y is separable and $x(.) \in \mathfrak{L}^\infty(S)[Y^*, Y]$, then* (i) *$x(.) \in \mathfrak{L}_0^\infty(S)[Y^*, Y]$,* (ii) *the mappings*

$$V(y) = x(s)(y), \qquad U(\phi') = \int_S x(s)\phi'(s)d\alpha,$$

where $\int_S x(s)\phi'(s)d\alpha$ is the Y-integral of $x(.)\phi'(.)$, are operations from $\bar{Y}$ to $L^\infty(S)$ and from $L(S)$ to Y^* respectively, (iii) *V coincides over Y with the adjoint of U, and* (iv) ess. sup. $\|x(s)\| = C < \infty$. *The finite constant C is the common norm of U, V, and the functional $F(y, \phi') = \int_S f(x(s))\phi'(s)d\alpha$ bilinear over $Y \times L(S)$.*

Conclusions (i)–(iii) are corollaries of Theorem 1.1.4. The remainder follows from Theorem 9 of [10].

C. **Measurable functions and absolutely integrable functions.** Beginning again from a somewhat different viewpoint, $x(s)$ is *measurable* [3] if it is the

limit a.e. of some sequence of simple functions; since S is decomposable this is equivalent to being the limit a.e. of a sequence of finitely-valued functions. For each p in the range $1 \leqq p < \infty$ there exists the subclass of measurable functions having $\int_S \|x(s)\|^p d\alpha < \infty$. Denoting this subclass by $\mathfrak{A}^p(S)[X]$ and letting $\mathfrak{A}^\infty(S)[X]$ be those measurable functions for which ess. sup. $\|x(s)\| < \infty$ it follows, as Garrett Birkhoff has shown [2, 9], that $\mathfrak{A}^p(S)[X] \subset \mathfrak{L}_0^p(S)[X, X^*]$ for $1 \leqq p \leqq \infty$. The elements of $\mathfrak{A}^1(S)[X]$ will be called *absolutely integrable.*

THEOREM 1.1.6. *If $x(.) \in \mathfrak{A}^p(S)[X]$, the function $x(.)\phi'(.)$ is absolutely integrable for each $\phi'(.) \in L^{p'}(S)$ and $U(\phi') = \int_S x(s)\phi'(s)d\alpha$ defines an operation U from $L^{p'}(S)$ to X. When $p = \infty$ the norm of U is $|U| =$ ess. sup. $\|x(s)\|$, and*

$$|U| \leqq \left(\int_S \|x(s)\|^p d\alpha\right)^{1/p}$$

when $p < \infty$.

The absolute integrability of $x(.)\phi'(.)$ should be evident. The existence and linearity of U are likewise obvious, in view of Theorem 1.1.3 and the inclusion of $\mathfrak{A}^p(S)[X]$ in $\mathfrak{L}_0^p(S)[X, X^*]$. For the case $p = \infty$ it is known [31, 3.11] that the operation $V(f) = f(x(.))$ from X^* to $L^\infty(S)$ has $|V| =$ ess. sup. $\|x(s)\|$; since V is the adjoint of U this means that $|U| =$ ess. sup. $\|x(s)\|$. The concluding inequality, for the case $p < \infty$, has already been proved in Theorem 2.4 of [9].

It is known that the two properties of weak measurability and almost separable valuedness together are sufficient as well as necessary for measurability [18, 31]. This remains true when a less stringent condition is substistituted for weak measurability; and if a third assumption is made, $x(.)$ becomes not only measurable but also essentially bounded.

THEOREM 1.1.7. *For a given $x(.)$ suppose that for some null set E_0 the set $x(S - E_0)$ has in X a separable span Y and that $x(.)$ is Γ-measurable for some manifold Γ which is determining for Y. Then* (i) *$x(.)$ is measurable. If it is also true that* ess. sup.$_s$ $|f(x(s))| < \infty$ *holds for each $f \in \Gamma$ and Γ is closed and linear, then* (ii) *$x(.) \in \mathfrak{A}^\infty(S)[X]$, and hence* (iii) *$x(.)\phi(.)$ is absolutely integrable for each $\phi \in L(S)$ and $U(\phi) = \int_S x(s)\phi(s)d\alpha$ defines as operation U from $L(S)$ to X having $|U| =$ ess. sup. $\|x(s)\|$.*

Conclusion (i) has been established in Theorem 2.1 of [32]. If it is also true that Γ is a c.l.m., then, since Y is separable, there is in Γ a separable c.l.m. Γ' which is determining for Y. Thus if $x(.) \in \mathfrak{L}^\infty(S)[X, \Gamma]$ clearly $x(.) \in \mathfrak{L}^\infty(S)[Y, \Gamma']$; from Theorem 9 of [10] it follows that ess. sup. $\|x(s)\|$

$<\infty$, so that the measurable function $x(.)$ is in $\mathfrak{L}^{\infty}(S)[X]$. Conclusion (iii) results from (ii) and Theorem 1.1.6.

PART 2. SOME CONVEX SETS RELATED TO INTEGRABLE FUNCTIONS

This part will consist of a few results involving inclusion relationships between certain closed convex sets containing the functional values of an integrable function and other closed convex sets associated with the function's indefinite integral. In a B-space X the closed convex hull of a subset Y will be denoted by $C[Y]$. For a fixed function $x(.)$ which is integrable over every $E \in \mathcal{E}_B$ the set of all quotients $(\int_E x(s)d\alpha)/\alpha(E)$ with $0<\alpha(E)<\infty$ will be represented by J. Thus J is defined if $x(.) \in \mathfrak{L}_0^p(S)[X, X^*]$, $1 \leqq p \leqq \infty$.

THEOREM 1.2.1. *If x_0 is a point disjoint with a closed convex set X', there exist an $f_0 \in X^*$ and a constant c such that $f_0(x_0)>c$ while $f_0(x) \leqq c$ for $x \in X'$.*

By a theorem of Mazur [27, p. 80] there is a closed convex body K disjoint with x_0 and containing X'. There must then be a nondegenerate closed sphere N about x_0 such that NK is vacuous. From a theorem of Eidelheit [12] the disjunction of the two closed convex bodies N and K implies the existence of an $f_0 \in X^*$ and a constant c such that $\|f_0\| \neq 0$, $f_0(x) \geqq c$ for $x \in N$, and $f_0(x) \leqq c$ for $x \in K$. Since $\|f_0\| \neq 0$ and no nondegenerate sphere can be contained in a hyperplane, there is an $x_1 \in N$ failing to satisfy the equation $f_0(x)=c$; thus $f_0(x_1)=c+\delta$ where $\delta>0$. If now $f_0(x_0)=c$, it follows that $f_0(2x_0-x_1)=c-\delta<c$ where $2x_0-x_1$ is in N. This contradicts the inequality $f_0(x) \geqq c$ for $x \in N$, and hence $f_0(x_0)>c$ must hold.

THEOREM 1.2.2. *If $x(.)$ is integrable over every $E \in \mathcal{E}_B$, then $C[J] \subset C[x(S-E_0)]$ for each $E_0 \in \mathcal{E}_0$.*

Let $x_E=\int_E x(s)d\alpha$, $E \in \mathcal{E}_B$. If there were an $E \in \mathcal{E}_B-\mathcal{E}_0$ for which $x_E/\alpha(E) \notin C[x(S-E_0)]$, then by the preceding theorem a constant c and an $f_0 \in X^*$ would exist such that $f_0(x_E/\alpha(E))>c$ while $x \in C[x(S-E_0)]$ implies $f_0(x) \leqq c$. Since $\alpha(E_0)=0$ we would then have $f_0(x(s)) \leqq c$ a.e. in E, so that $\int_E f_0(x(s))d\alpha \leqq c\alpha(E)$. Yet $\int_E f_0(x(s))d\alpha=f_0(x_E)>c\alpha(E)$, which is a contradiction. Thus $J \subset C[x(S-E_0)]$ and hence $C[J] \subset C[x(S-E_0)]$.

An immediate consequence of Theorem 1.2.2 is

THEOREM 1.2.3. *For a function $x(.)$ integrable over every $E \in \mathcal{E}_B$ the inclusion $C[J] \subset \prod C[x(S-E_0)]$ holds, the product being taken as E_0 varies over $\mathcal{E}_0$.*

Another easy corollary is given in

THEOREM 1.2.4. *If $x(.) \in \mathfrak{L}_0^p(S)[X, X^*]$ and Y is a c.l.m. containing $x(S-E_0)$ for some $E_0 \in \mathcal{E}_0$ then $Y \ni U(\phi')=\int_S x(s)\phi'(s)d\alpha$ for every $\phi' \in L^{p'}(S)$,*

that is, $Y \supset U(L^{p'}(S))$. Thus U is a separable operation when $x(.)$ is almost separably-valued.

Since Y is a closed convex set containing $x(S-E_0)$, Theorem 1.2.2 implies that $Y \supset J$. From the linearity of Y it then follows that $Y \ni \int_E x(s)d\alpha$ for each $E \in \mathcal{E}_B$. If $p=1$ the complete additivity of $\int_E x(s)d\alpha$ over $\mathcal{E}$ [31, 10] and the fact that Y is closed and linear yield the conclusion that $Y \ni \int_E x(s)d\alpha$ for every $E \in \mathcal{E}$. The simple functions being dense in $L^{p'}(S)$ for $p' < \infty$ and the finitely-valued ones being dense in $L^{\infty}(S)$, it follows from the linearity of U that $U(\phi')$ is in the c.l.m. Y for each $\phi' \in L^{p'}(S)$.

COROLLARY. *If $x(.) \in \mathfrak{A}^p(S)[X]$, then $U(\phi') = \int_S x(s)\phi'(s)d\alpha$ is a separable operation from $L^{p'}(S)$ to X.*

Theorem 1.2.4 can be given a considerably sharper form when $p=\infty$. For this purpose let $R(x) = -x$ for $x \in X$, and let $\mathfrak{P}$ be the essentially non-negative elements of $L(S)$. The subset of $\mathfrak{P}$ for which $\|\phi\| = 1$ holds will be denoted by $\mathfrak{P}_1$.

THEOREM 1.2.5. *Suppose that $x(.)$ is in $\mathfrak{L}_0^{\infty}(S)[X, X^*]$ and that Y is a closed convex set containing $x(S-E_0)$ for some $E_0 \in \mathcal{E}_0$. For $U(\phi) = \int_S x(s)\phi(s)d\alpha$, $\phi \in L(S)$, it follows that*

(i) *$U(\phi) \in C[x(S-E_0)] \subset Y$ when $\phi \in \mathfrak{P}_1$,*
(ii) *if $Y \ni \theta$ then $U(\phi) \in Y$ for each $\phi \in \mathfrak{P}$ having $\|\phi\| \leqq 1$,*
(iii) *if $Y \supset R(x(S-E_0))$ then $U(\phi) \in Y$ for each ϕ with $\|\phi\| \leqq 1$,*
(iv) *if Y is linear then $U(\phi) \in Y$ for every ϕ.*

In (i) the second relationship is obvious. To obtain the first it is sufficient to show that $Y \ni U(\phi)$ whenever ϕ is an essentially non-negative simple function having $\|\phi\| = 1$. For such a ϕ we have $U(\phi) = \sum_1^n \phi_i \int_{E_i} x(s)d\alpha = \sum_1^n \alpha_i x_i$ where $\alpha_i \equiv \phi_i \alpha(E_i) \geqq 0$, $\infty > \alpha(E_i) > 0$, and $x_i = 1/\alpha(E_i) \cdot \int_{E_i} x(s)d\alpha$. Since $x_i \in J$, $\alpha_i \geqq 0$, and $\sum_1^n \alpha = \|\phi\| = 1$, it follows that $U(\phi) \in C[J]$ and hence (Theorem 1.2.2) $U(\phi) \in C[x(S-E_0)]$. Statement (ii) results immediately from (i), the homogeneity of U, and the convexity of Y. Conclusion (iii) is also implied by (i). For an arbitrary $\phi \in L(S)$ can be written as $\phi = \phi_1 - \phi_2$ where $\phi_i \in \mathfrak{P}$ and $\|\phi\| = \sum_1^2 \|\phi_i\|$. Thus

$$U(\phi/\|\phi\|) = \sum_1^2 \rho_i x_i, \qquad \|\phi\| > 0,$$

where $\rho_i = \|\phi_i\|/\|\phi\| \geqq 0$, $x_1 = U(\phi_1/\|\phi_1\|)$, and $x_2 = -U(\phi_2/\|\phi_2\|)$. Since (i) implies that $x_1 \in C[x(S-E_0)]$ and $x_2 \in R(C[x(S-E_0)])$, and since $\rho_1 + \rho_2 = 1$ where $\rho_i \geqq 0$, it is clear that $Y \ni U(\phi/\|\phi\|)$ due to the convexity of Y. We can then conclude, Y being a convex set containing θ, that $Y \ni U(\phi)$ for

every ϕ with $\|\phi\| \leqq 1$. The final statement, (iv), results from either (iii) or Theorem 1.2.4.

THEOREM 1.2.6. *Suppose that for* $x(.) \in \mathfrak{L}_0^\infty(S)[X, X^*]$ *there exist a null set* E_0, *a closed convex set* Y, *and a constant* C *such that* $Y \supset x(S-E_0)$ *and* $\|y\| = C$ *for each* $y \in Y$. *Then* $U(\phi) \in Y$ *for every* $\phi \in \mathfrak{P}_1$ *and hence* $\|U(\phi)\| = C\|\phi\|$ *for every* $\phi \in \mathfrak{P}$.

It will now be shown that for a restricted $x(s)$ the inclusion reverse to that of Theorem 1.2.2 holds for at least one $E_0 \in \mathcal{E}_0$.

THEOREM 1.2.7. *If* $x(.)$ *is integrable over every* $E \in \mathcal{E}_B$ *and is also measurable, there is an* $E_0' \in \mathcal{E}_0$ *such that* $C[x(S-E_0')] \subset C[J]$.

Since $x(.)$ is measurable, there is a null set E'' such that $x(S-E'')$ is separable. Thus if $x(.)$ is redefined so as to vanish over E'', the new function $x'(.)$ will be measurable and $x_E' \equiv \int_E x'(s)d\alpha = \int_E x(s)d\alpha$ for every $E \in \mathcal{E}_B$. Moreover $x'(S)$ will be a separable set in X. If in $\mathcal{E}_0$ there is an E' with the property that $C[x'(S-E')] \subset C[J']$, it will then follow, should the theorem be true for $x'(.)$, that $C[x(S-E''-E')] \subset C[x'(S-E')] \subset C[J'] = C[J]$, where $E_0' = E''+E' \in \mathcal{E}_0$. Hence we may suppose that $x(S)$ is separable. From the measurability of $x(.)$ it can be concluded that $x^{-1}(Y)$ is in $\mathcal{E}$ for any open sphere Y in X; since $x(S)$ is a separable set in X this remains true if Y is an arbitrary open set and hence if Y is any Borel-measurable set. It is now evident, $C[J]$ being closed, that the set $E' = x^{-1}(X-C[J])$ must be measurable. To prove the theorem we have only to show that $\alpha(E') = 0$.

Each point $x_0 \in x(E')$ is disjoint with $C[J]$; to each such x_0 there corresponds by Theorem 1.2.1 an element f of X^* and a constant c such that $f(x_0) > c$ and $f(x) \leqq c$ for $x \in C[J]$. From the continuity of f an open sphere N about x_0 exists with the property that $f(x) > c$ holds for every $x \in N$. Let each point of $x(E')$ be covered by a sphere of this kind. Since $x(E')$ is a separable set in X, a denumerable number N_m, $m = 1, 2, \cdots$, of these spheres suffice to cover $x(E')$. Letting $E_m = x^{-1}(N_m)$ it is clear that E_m is measurable and that $\sum_1^\infty E_m \supset E'$. Thus if $\alpha(E')$ is positive, then $\alpha(E_{m_0}) > 0$ holds for some m_0. Taking a decomposition $\{S_n\}$ of S, the inequality $\infty > \alpha(S_n E_{m_0}) > 0$ must be true for at least one n. Letting E^* be the set $S_n E_{m_0}$ for one such n, obviously $0 < \alpha(E^*) < \infty$ and $x_{E^*}/\alpha(E^*) \in J$. From the definitions of N_{m_0}, E_{m_0}, and E^* there is an $f_0 \in X^*$ and a constant c_0 such that $f_0(x(s)) > c$ for every $s \in E^*$, while $f_0(x) \leqq c_0$ when $x \in C[J]$. Due to the inequality $\infty > \alpha(E^*) > 0$ we then have $\int_{E^*} f_0(x(s))d\alpha > c_0\alpha(E^*)$ and at the same time $\int_{E^*} f_0(x(s))d\alpha = f_0(x_{E^*}) \leqq c_0\alpha(E^*)$. This contradiction ends the proof since it implies $\alpha(E') = 0$.

Combining Theorems 1.2.3 and 1.2.7 gives us

THEOREM 1.2.8. *If $x(.)$ is integrable over every $E \in \mathcal{E}_B$ and is measurable, then an $E_0' \in \mathcal{E}_0$ exists such that*

$$C[x(S - E_0')] = C[J] = \prod C[x(S - E_0)],$$

the product being taken as E_0 varies over $\mathcal{E}_0$.

For measurable functions Theorem 1.2.4 may be rounded out as follows.

THEOREM 1.2.9. *Let $x(.) \in \mathfrak{L}_0^p(S)[X, X^*]$ be measurable. For a c.l.m. Y the following are equivalent*: (i) *$x(.)$ is essentially defined to Y*, (ii) *$Y \supset \int_E x(s)d\alpha$ for every $E \in \mathcal{E}_B$, and* (iii) *$Y \supset U(\phi) = \int_S x(s)\phi(s)d\alpha$ for every $\phi \in L^{p'}(S)$.*

For (iii) clearly implies (ii), and (iii) follows from (i) by Theorem 1.2.4. If (ii) holds, then $Y \supset C[J]$, and hence $Y \supset x(S - E_0')$ for some $E_0' \in \mathcal{E}_0$ according to Theorem 1.2.8.

When $p = \infty$ more precise results can be obtained.

THEOREM 1.2.10. *Suppose $x(.)$ in $\mathfrak{L}_0^\infty(S)[X, X^*]$ is measurable. Then* (I) *an $E_0' \in \mathcal{E}_0$ exists such that*

$$\prod_{E_0 \in \mathcal{E}_0} C[x(S - E_0)] = C[x(S - E_0')] = C[J] = C[U(\mathfrak{P}_1)].$$

Let Y be a closed convex set. Then (II) *these three conditions are equivalent*:

(i) *$Y \supset x(S - E_0)$ for some $E_0 \in \mathcal{E}_0$,*

(ii) *$Y \supset 1/\alpha(E) \cdot \int_E x(s)d\alpha$ for every $E \in \mathcal{E}_B - \mathcal{E}_0$,*

(iii) *$Y \supset U(\phi) = \int_S x(s)\phi(s)d\alpha$ for every $\phi \in \mathfrak{P}_1$;*

(III) *these three are equivalent*:

(i) *Y satisfies* (II)(i) *and $Y \supset \theta$,*

(ii) *Y satisfies* (II)(ii) *and $Y \supset \theta$,*

(iii) *$Y \supset U(\phi)$ for every $\phi \in \mathfrak{P}$ with $\|\phi\| \leqq 1$;*

and (IV) *these are equivalent*:

(i) *for some $E_0 \in \mathcal{E}_0$ both $x(S - E_0)$ and $R(x(S - E_0))$ are in Y,*

(ii) *$Y \supset (\delta/\alpha(E))\int_E x(s)d\alpha$ for every $E \in \mathcal{E}_B - \mathcal{E}_0$ and every $\delta = \pm 1$,*

(iii) *$Y \supset U(\phi)$ for each ϕ with $\|\phi\| \leqq 1$.*

(V) *Finally, for a given constant K these are equivalent*:

(i) *$\|U(\phi)\| = K$ for every $\phi \in \mathfrak{P}_1$,*

(ii) *an $E_0' \in \mathcal{E}_0$ exists such that $\|x\| = K$ for $x \in C[x(S - E_0')]$.*

Hence if (V)(i) *or* (V)(ii) *holds, then $\|x(s)\| = K$ a.e. and $\|\int_E x(s)d\alpha\| = K\alpha(E)$ for every $E \in \mathcal{E}_B$.*

(I) follows immediately from Theorem 1.2.8, conclusion (i) of Theorem 1.2.5, and the fact that $J \subset U(\mathfrak{P}_1)$. (II) and (III) are evident consequences of (I) and (II) respectively. (IV) results from (III) and Theorem 1.2.5 on considering the operation $U'(\phi) = \int_S (-x(s))\phi(s)d\alpha = -U(\phi)$. The equiva-

lence of the two conditions in (V) follows from (I) and the fact that $C[U(\mathfrak{P}_1)]$ is simply the closure of the convex set $U(\mathfrak{P}_1)$.

Example 3 in Birkhoff's paper [2] shows that Theorems 1.2.7–1.2.10 all become false if the hypothesis of measurability for $x(.)$ be dropped.

PART 3. KERNEL REPRESENTATIONS OF ABSTRACT FUNCTIONS AND INTEGRALS

We now consider a second system $[T, \mathcal{T}, \beta]$ entirely analogous to $[S, \mathcal{E}, \alpha]$, with $\mathcal{T}_0$ denoting all those elements of $\mathcal{T}$ having β-measure zero and $\mathcal{T}_B$ all those having finite β-measure.

A. **Measurable functions in terms of numerical kernels.** When the range space X consists of numerical functions defined and measurable over T, it sometimes happens that each measurable function $x(s)$ having its values in X can be represented by means of a numerical kernel defined and measurable* over $S \times T$. The first theorem below gives a set of conditions on X that are sufficient for the existence of such representations.

THEOREM 1.3.1. *Suppose X satisfies these conditions*:

(a) *if $\psi^m \in X$, $m=1, 2, \cdots$, and $\lim_m \|\psi^m\| = 0$, then $\{\psi^m(t)\}$ converges to zero in measure over every $F \in \mathcal{T}_B$,*

(b) *$\psi(.) \in X$ implies $\|\psi\| = 0$ if and only if $\psi(t) = 0$ a.e.,*

(c) *if $\{T_i\}$ is a decomposition of T and with each $\psi(.) \in X$ are associated the functions $\psi_n(t) = \psi(t)\phi_n(t)$, $n=1, 2, \cdots$, where $\phi_n(t)$ is the characteristic function of $\sum_1^n T_i$, then*

(i) *$\psi_n(.) \in X$ for each n,*

(ii) *$\lim_m \psi^m = \psi$ in X implies $\lim_m \psi_n^m = \psi_n$ in X for each n.*

Under these conditions it follows that if x_s is measurable from S to X there is a numerical kernel $K(s, t)$ measurable over $S \times T$ and such that for almost† every $s \in S$ the equality $x_s(t) = K(s, t)$ holds a.e. in T.‡ Hence if (b) *is strengthened to*

(b′) *$\psi(t)$ is measurable and $\psi(t) = 0$ almost everywhere if and only if $\psi(.) \in X$ and $\|\psi\| = 0$,*

then for almost every s $K(s, .)$ is in X and coincides with the point x_s.

The demonstration of this theorem for the case $\alpha(S) < \infty$ and $\beta(T) < \infty$ has been given in 3.1 of [9].§ The extension to the case of a general decomposable T follows from a reapplication of the same methods. The final step, to a general decomposable S, is then easily made.

* That is, measurable with respect to the Borel field of subsets of $S \times T$ determined by the two fields E and F [33].

† If x_s is defined over all of S the word "almost" may be deleted.

‡ $K(s, t)$ is then said to be a *measurable representation of $x(s)$*.

§ The proof given there is for Euclidean S and T but can be carried over to the present case without change.

Since $L^q(T)$ satisfies conditions (a), (b'), and (c) for $1 \leqq q \leqq \infty$ we have [9]

THEOREM 1.3.2. *If $x(s)$ is measurable from S to $L^q(T)$, a measurable kernel $K(s, t)$ exists over $S \times T$ such that $K(s, .) = x(s)$ in $L^q(T)$ a.e. in S. Any two such representations of $x(.)$ differ over $S \times T$ on at most a set of measure zero.*

On the other hand the converse (partial when $q = \infty$) of Theorem 1.3.2 also holds.

THEOREM 1.3.3. *Let $K(s, t)$ be a kernel over $S \times T$ and let q be fixed. Suppose that*

(i) *$K(s, t)$ is measurable,*

(ii) *$K(s, .) \in L^q(T)$ for almost every s, and*

(iii) *if $q = \infty$ then $K(s, .)$ is almost separably-valued in $L^q(T)$.*

Then $x(s) = K(s, .)$ as a function defined to $L^q(T)$ is almost separably-valued and is measurable.

We first show that $K(s, .)$ is almost separably-valued when $q < \infty$. Suppose $K(s, t)$ is the characteristic function of a set having finite measure in the product space $S \times T$. There is then a sequence $\{K_n(s, t)\}$ such that each $K_n(s, t)$ is the characteristic function of a finite sum of sets of the form $E \times F$ and

$$\lim_n \iint_{S \times T} | K_n(s, t) - K(s, t) |^q d(\alpha \times \beta) = 0.$$

Thus there is a subsequence $\{K_{n_i}\}$ such that

$$\lim_i \int_T | K_{n_i}(s, t) - K(s, t) |^q d\beta = 0$$

a.e. in S. This shows that the values of $x(s) = K(s, .)$ lie for almost all s in a separable subset of $L^q(T)$. If K is the characteristic function of a set of infinite measure in the product space, then $K = \sum_n \phi_n K$ where ϕ_n is the characteristic function of G_n and the sequence $\{G_n\}$ forms a partition of $S \times T$ into sets of finite measure. So in this case also we have $x(.)$ almost separably-valued. Thus the abstract function $x(s) = K(s, .)$ corresponding to a finitely-valued kernel $K(s, t)$ is almost separably-valued. Now suppose $K(s, t)$ is an arbitrary function satisfying (i) and (ii) for $q < \infty$. Since K is measurable on $S \times T$ there is a sequence $\{K_n(s, t)\}$ of finitely-valued functions having the properties

$$\lim_n K_n(s, t) = K(s, t) \qquad \text{a.e. on } S \times T, \tag{1}$$

$$| K_n(s, t) - K(s, t) | \leqq | K(s, t) | \qquad \text{on } S \times T. \tag{2}$$

The Fubini theorem (1) shows that, for almost every s, $\lim_n K_n(s, t) = K(s, t)$ a.e. in T. This fact together with (2), (ii), and the Lebesgue convergence theorem yields for almost every s

$$\lim_n \int_T | K_n(s, t) - K(s, t) |^q d\beta = 0,$$

which proves that $x(s) = K(s, .)$ is almost separably-valued.

Thus $x(s) = K(s, .)$ is almost separably-valued for $1 \leqq q \leqq \infty$. Since the measurability of K implies that $\int_T K(s, t)\psi'(t)d\beta$ is measurable in s for each $\psi'(.) \in L^{q'}(T)$, it then follows from Theorem 1.1.7 that $x(.)$ is measurable.

B. **The representation of abstract integrals by kernel integrals.** We recall the following two theorems giving representations for certain integrals. These have been obtained previously (for Euclidean S and T) in [31] and [9].

THEOREM 1.3.4. *Suppose $x(.)$ in $\mathfrak{L}_0^{p'}(S)[L^q(T), L^{q'}(T)]$ has a measurable representation $K(s, t)$. A necessary and sufficient condition that the operation* $U(\phi) = \int_S x(s)\phi(s)d\alpha$ from $L^p(S)$ to $L^q(T)$ be expressible as*

$$U(\phi) = \int_S K(s, t)\phi(s)d\alpha$$

is that $\phi \in L^p(S)$ and $\psi' \in L^{q'}(T)$ imply that $\int_T \psi'(t)\{\int_S K(s, t)\phi(s)d\alpha\}d\beta = \int_S \phi(s)\{\int_T K(s, t)\psi'(t)d\beta\}d\alpha$ finitely.

This is merely Theorem 7.3 of [31] carried over to the present more general S and T; the proof remains precisely the same.

From Theorems 1.3.2 and 1.3.4 we can now derive

THEOREM 1.3.5 [9]. *Suppose $x(.)$ is in $\mathfrak{L}^{p'}(S)[L^q(T)]$. Then a kernel $K(s, t)$ exists with these properties:*

(i) *$K(s, t)$ is measurable, and*

(ii) *$K(s, .) = x(s)$ in $L^q(T)$ for almost every s.*

If $K(s, t)$ is any kernel satisfying (i) *and* (ii), *then*

(iii) $(\int_S \|K(s, .)\|^{r'} d\alpha)^{1/r'} \equiv C < \infty$ *when* $p' < \infty$ *and* ess. sup. $\|K(s, .)\| \equiv C < \infty$ *when* $p' = \infty$,

(iv) *the separable operation $U(\phi) = \int_S x(s)\phi(s)d\alpha$ can be written as $U(\phi) = \int_S K(s, t)\phi(s)d\alpha$, $\phi \in L^p(S)$,*

(v) $|U| \leqq C$ *for* $p' < \infty$ *and* $|U| = C$ *for* $p' = \infty$.

The existence of a $K(s, t)$ satisfying (i) and (ii) comes immediately from Theorem 1.3.2. Property (iii) is obvious in view of (ii) and the fact that $x(.) \in \mathfrak{L}^{p'}(S)[L^q(T)]$. In (iv) the operation U exists and is separable by Theo-

* When $q = \infty$, this integral is an $L(T)$-integral.

rem 1.1.6 and the corollary to Theorem 1.2.4. Moreover if $K(s, t)$ satisfies (i) and (ii) we have for each $\psi'(.) \in L^{q'}(T)$ that

$$\int_T |K(s, t)| \cdot |\psi'(t)| \, d\beta \leqq \|\psi'\| \cdot \|x(s)\|,$$

by Hölder's inequality. Since $\|x(.)\| \in L^{p'}(S)$ this means that $\int_S |\phi(s)| \cdot \{\int_T |K(s, t)| \cdot |\psi'(t)| \, d\beta\} d\alpha$ exists finitely whenever $\phi(.) \in L^p(S)$ and $\psi'(.) \in L^{q'}(T)$. Since $K(s, t)\phi(s)\psi'(t)$ is measurable, a classical theorem of Tonelli implies that the two repeated integrals in Theorem 1.3.4 exist finitely and are equal. From the latter theorem (iv) then follows. Finally, (v) results from (ii) and Theorem 1.1.6.

If in Theorem 1.3.5 we begin with a kernel instead of an abstract function, we can state

THEOREM 1.3.6. *Let $K(s, t)$ be a kernel satisfying for a fixed q conditions* (i)–(iii) *of Theorem* 1.3.3 *and for a fixed p' condition* (iii) *of Theorem* 1.3.5. *Then $x(s) = K(s, .)$ is in $\mathfrak{A}^{p'}(S)[L^q(T)]$ and $U(\phi) = \int_S K(s, t)\phi(s)d\alpha$ is an operation defined and separable from $L^p(S)$ to $L^q(T)$, with $|U| \leqq C$ when $p' < \infty$ and $|U| = C$ when $p' = \infty$.*

It is clear from Theorem 1.3.3 and property (iii) of Theorem 1.3.5 that $x(s) = K(s, .)$ is in $\mathfrak{A}^{p'}(S)[L^q(T)]$. The remaining conclusions follow from Theorem 1.3.5.

Later it will be seen that the operation U in Theorem 1.3.6 is c.c. when $p' < \infty$ [38] (Theorems 3.2.11 and 3.2.12) and that U takes weakly compact sets in $L(S)$ into compact sets in $L^q(T)$ when $p' = \infty$ (Theorem 3.2.1).

For the case $p' = \infty$ it results from Theorems 1.3.3, 1.1.7, and 1.3.6 that the fourth assumption in the last theorem can be weakened as follows.

THEOREM 1.3.7. *Suppose $K(s, t)$ satisfies conditions* (i)–(iii) *in Theorem* 1.3.3 *and in addition* ess. $\sup_s \left|\int_T K(s, t)\psi'(t)d\beta\right| < \infty$ *is true for each $\psi'(.) \in L^{q'}(T)$. Then $x(s) = K(s, .)$ is in $\mathfrak{A}^{\infty}(S)[L^q(T)]$ and the operation $U(\phi) = \int_S K(s, t)\phi(s)d\alpha$ is defined and separable from $L(S)$ to $L^q(T)$ with $|U| = \text{ess. sup.}_s \left(\int_T |K(s, t)|^q d\beta\right)^{1/q}$ for $q < \infty$ and $|U| = \text{ess. sup.}_{s,t} |K(s, t)|$ for $q = \infty$.*

CHAPTER II. GENERAL OPERATIONS TO RESTRICTED SPACES

PART 1. THE REPRESENTATION IN ADJOINT SPACES OF SET FUNCTIONS AND LINEAR OPERATIONS BY ABSTRACT INTEGRALS

A. Integral representations of absolutely continuous and Lipschitzean set functions. The first theorem below may be regarded either as a generalization (to the case of an abstract S) of a result of Gelfand [18] or as an extension

of the Radon-Nikodym theorem [29, 34] from real-valued set functions to those taking their values in an adjoint space. Two proofs of the theorem will be given.

THEOREM 2.1.0. *Let $X=[x]$ be the adjoint of another B-space $Y=[y]$ and let x_E be defined from $\mathcal{E}_B$ to X. Suppose that Y' is a separable linear subset of Y and that*

(i) *for each $y \in Y'$ the real-valued set function $x_E(y)$ is completely additive over* $\mathcal{E}(E')$ for every $E' \in \mathcal{E}_B$,*

(ii) *$x_E(y)=0$ when E is a null set and $y \in Y'$, and*

(iii) *the numerical function*

$$\sigma_E \equiv \sup_{y} \frac{1}{\|y\|} |\, x_E(y)\,|, \qquad y \in Y',\ y \neq \theta,$$

which is defined over $\mathcal{E}_B$ has $\nu_{E'}$, its total variation over E', finite for every $E' \in \mathcal{E}_B$.

Then there exists an x_s defined from S to X such that

$$x_E(y) = \int_E x_s(y)\,d\alpha, \qquad E \in \mathcal{E}_B,\ y \in Y', \tag{2.1.01}$$

$$\nu_E = \int_E \|x_s\|\,d\alpha \qquad \text{for every } E \in \mathcal{E}_B. \tag{2.1.02}$$

First proof. The initial step† is to show that if $E \in \mathcal{E}_B$ and $\{E_j\}$ is a finite or denumerable number of disjoint elements of $\mathcal{E}$ with $\sum_j E_j = E$ then

$$\nu_E = \sum_j \nu_{E_j}. \tag{1}$$

Since it is obvious from the definition of ν_E that $\nu_E \geqq \sum_j \nu_{E_j}$, it will be sufficient to establish the reverse inequality

$$\nu_E \leqq \sum_j \nu_{E_j}. \tag{2}$$

If we choose disjoint measurable sets $E_1', \cdots, E_n'$ in E such that $\nu_E - \epsilon < \sum_1^n \sigma_{E_i'}$ where $\epsilon > 0$, clearly there are elements $y_1, \cdots, y_n$ in Y' for which

$$\nu_E - \epsilon < \sum_{i=1}^{n} \rho_i\, |\, x_{E_i'}(y_i)\,|,$$

where $\rho_i \|y_i\| = 1$. Since (i) implies that $x_{E_i'}(y_i) = \sum_j x_{E_i' E_j}(y_i)$ and $\sum_j |\, x_{E_i' E_j}(y_i)\,|$

* For a fixed $E' \in \mathcal{E}$ the symbol $\mathcal{E}(E')$ denotes the Borel field composed of all sets of the form EE' where $E \in \mathcal{E}$.

† The theorem is trivial when Y' is vacuous or consists of the zero element.

$< \infty$ for each i, we can conclude that

$$\nu_E - \epsilon < \sum_{i=1}^{n} \rho_i \left| \sum_j x_{E_i' E_j}(y_i) \right| < \sum_j \sum_i \rho_i \left| x_{E_i' E_j}(y_i) \right|$$

and hence that $\nu_E - \epsilon < \sum_i \sum_i \sigma_{E_i' E_j} \leqq \sum_i \nu_{E_j}$ whenever $\epsilon > 0$. This vindicates (2), so that (1) holds and ν_E is completely additive over $\mathcal{E}(E')$ for each $E' \in \mathcal{E}_B$. An additional property of ν_E is that $\nu_E = 0$ when E is a null set, by virtue of (ii). In view of the decomposability of S and the Radon-Nikodym theorem these two properties lead to the existence of a measurable non-negative function $\phi_0(s)$ satisfying the condition that

$$\int_{E'} \phi_0(s) d\alpha = \nu_{E'} \qquad \text{for every } E' \in \mathcal{E}_B. \tag{3}$$

S being decomposable it is also true that

$$|\phi_0(s)| < \infty \qquad \text{a.e. in } S. \tag{4}$$

Let $\{Y_m\}$ be a denumerable dense subset of Y'. Due to (i) and (ii) the decomposability of S and the Radon-Nikodym theorem also imply that for each m a measurable real function $\phi_m(s)$ exists such that

$$x_E(y_m) = \int_E \phi_m(s) d\alpha, \qquad E \in \mathcal{E}_B. \tag{5}$$

Denoting by Y_0 the set of all finite rational linear combinations of the y_m's we then have

$$x_E(y) = \sum_1^n h_j x_E(y_j) = \int_E \sum_1^n h_j \phi_j(s) d\alpha, \qquad E \in \mathcal{E}_B,$$

whenever $y \in Y_0$ and $y = \sum_1^n h_j y_j$. Hence for $y \in Y_0$ and $E \in \mathcal{E}_B$ the total variation of $x_E(y)$ over E' is $\int_{E'} \left| \sum_1^n h_j \phi_j(s) \right| d\alpha$. But $Y_0 \subset Y'$ since Y' is linear· thus the total variation of $x_E(y)$ over E' is not greater than that of $\|y\| \sigma_E$ when $y \in Y_0$. If y has the form $y = \sum_1^n h_j y_j$ where each h_j is rational, then we have

$$\int_{E'} \left| \sum_1^n h_j \phi_j(s) \right| d\alpha \leqq \|y\| \nu_{E'} = \int_{E'} \|y\| \phi_0(s) d\alpha, \qquad E' \in \mathcal{E}_B,$$

whence

$$\left| \sum_1^n h_j \phi_j(s) \right| \leqq \phi_0(s) \|y\| \qquad \text{a.e. in } S \tag{6}$$

for each such y. The set Y_0 being denumerable, (6) and (4) yield a null set E_0 such that

$$\left| \sum_1^n h_j \phi_j(s) \right| \leqq \phi_0(s) \left\| \sum_1^n h_j y_j \right\| < \infty \tag{7}$$

holds for every $s \in S - E_0$ and every set $h_1, \cdots, h_n$ of rational numbers. Now let $h_1, \cdots, h_m$ be any real numbers. If we choose rationals h_j^i such that $\lim_i h_j^i = h_j$ for $j = 1, 2, \cdots, m$, (7) implies that

$$\left| \sum_1^m h_j \phi_j(s) \right| = \lim_i \left| \sum_{j=1}^m h_j^i \phi_j(s) \right|, \qquad s \in S - E_0,$$

and that

$$\left| \sum_{j=1}^m h_j^i \phi_j(s) \right| \leqq \phi_0(s) \left| \sum_{j=1}^m h_j^i y_j \right| < \infty, \qquad s \in S - E_0,\ i = 1, 2, \cdots.$$

Hence

$$\left| \sum_1^m h_j \phi_j(s) \right| \leqq \phi_0(s) \left\| \sum_1^m h_j y_j \right\| < \infty \tag{8}$$

holds for every $s \in S - E_0$ and every finite set $h_1, \cdots, h_m$ of real numbers.

In view of (8), a well known theorem on moments [1, p. 57] leads to the existence, for each $s \in S - E_0$, of an element x_s in the adjoint X of Y' such that

$$\|x_s\| \leqq \phi_0(s), \qquad x_s(y_m) = \phi_m(s), \qquad \text{for } m = 1, 2, \cdots. \tag{9}$$

Over E_0 let x_s be the zero element of X. Due to the Hahn-Banach theorem on the extension of linear functionals, it may be supposed that not only (9) holds but also that

$$\|x_s\| = \sup_y |x_s(y)|, \qquad y \in Y', \|y\| = 1, \tag{10}$$

for every s. This will be used in establishing (2.1.02). To obtain (2.1.01) let y be arbitrary in Y' and choose a subsequence $\{y_{m_i}\}$ in $\{y_m\}$ with $\lim_i y_{m_i} = y$. Then $\lim_i x_s(y_{m_i}) = x_s(y)$ for every s and $\{x_s(y_{m_i})\}$ is a sequence of measurable functions with

$$|x_s(y_{m_i})| \leqq \phi_0(s) \sup_i \|y_{m_i}\|, \qquad s \in S - E_0,\ i = 1, 2, \cdots.$$

Applying Lebesgue's convergence theorem shows that $x_s(y)$ is summable over every $E \in \mathcal{E}_B$ and that

$$\int_E x_s(y)d\alpha = \lim_i \int_E x_s(y_{m_i})d\alpha, \qquad E \in \mathcal{E}_B. \tag{11}$$

From (5) and (9) we also have

$$x_E(y) = \lim_i x_E(y_{m_i}) = \lim_i \int_E \phi_{m_i}(s)d\alpha = \lim_i \int_E x_s(y_{m_i})d\alpha \tag{12}$$

for $E \in \mathcal{E}_B$. Relation (2.1.01) now follows by combining (11) and (12). To prove (2.1.02) let $\{z_n\}$ be denumerable and dense in the surface of the unit sphere of Y'. We then have from (10)

$$\|x_s\| = \limsup_n |x_s(z_n)|, \qquad s \in S, \tag{13}$$

so that $\|x_s\|$ is measurable and therefore summable over every $E \in \mathcal{E}_B$. From (iii) and (10) it is also evident that $\sigma_E = \limsup_n |x_E(z_n)|$ for every $E \in \mathcal{E}_B$. Relation (2.1.01) then implies that $\sigma_E = \limsup_n \left|\int_E x_s(z_n)d\alpha\right|$ and hence $\sigma_E \leqq \int_E \limsup_n |x_s(z_n)|\,d\alpha = \int_E \|x_s\|d\alpha \leqq \int_E \phi_0(s)d\alpha = \nu_E$ by (3) and (9), so that (2.1.02) is true.

Second proof. We first cite the following

LEMMA. (Doob.) *Let Y be a metric space and Y_0 a denumerable set in Y which has the point θ in Y as a point of accumulation. Let $K(y, s)$ be a real function defined for $y \in Y_0$ and $s \in S$ which is measurable in s for each $y \in Y_0$. Suppose further that for any sequence $\{Y_i\}$ of points in Y_0 with $\lim_i y_i = \theta$ we have $\lim K(y_i, s) = 0$ a.e. in S. Then there is a null set E_0 such that*

$$\lim_{y \to \theta,\, y \in Y_0} K(y, s) = 0, \qquad s \in S - E_0.$$

In a slightly different form this has been proved by Doob [7, Lemma 2, p. 758], so that we omit the proof here.

If x_E satisfies the assumptions of the theorem, it follows from the decomposability of S and the Radon-Nikodym theorem that there is a real function $K(y, s)$ defined for $y \in Y'$ and $s \in S$ which is measurable in s for each y and is such that

$$x_E(y) = \int_E K(y, s)d\alpha, \qquad y \in Y', E \in \mathcal{E}_B. \tag{14}$$

As in the first proof a measurable non-negative $\phi_0(s)$ exists satisfying (3) and (4). Since (14) implies that over each $E' \in \mathcal{E}_B$ the total variation of $x_E(y)$ is $\int_{E'} |K(y, s)|\,d\alpha$ and since this total variation is not greater than that of $\|y\|\sigma_E$, we then have

$$\int_{E'} |K(y, s)|\, d\alpha \leqq \|y\|\nu_{E'} = \|y\| \cdot \int_{E'} \phi_0(s) d\alpha, \qquad y \in Y', E' \in \mathcal{E}_B,$$

and hence

$$|K(y, s)| \leqq \|y\|\phi_0(s) \qquad \text{a.e. in } S \text{ for each } y \in Y'. \tag{15}$$

From (14) it follows that

$$K(ay + by', s) = aK(y, s) + bK(y', s) \tag{16}$$

holds a.e. in S. The set of points in S where (16) fails to hold may of course vary with the real numbers a and b as well as with the vectors y and y'. Let Y_0 consist of all finite linear combinations, with rational coefficients, of the elements of some sequence dense in the span Γ of Y'. Since Y' is linear we may take Y_0 to be in Y'. Due to (16) there is a null set E_0' such that

$$K(ay + by', s) = aK(y, s) + bK(y', s) \tag{17}$$

holds for y and y' in Y_0, a and b rational, and $s \in S - E_0'$. In view of (4), (15), and the inclusion of Y_0 in Y' we may suppose E_0' so chosen that

$$|K(y, s)| \leqq \|y\|\phi_0(s) < \infty, \qquad y \in Y_0, s \in S - E_0'. \tag{18}$$

Now using (18) together with the lemma of Doob we see that there is a null set E_0 containing E_0' and such that

$$\lim_{y \to \theta,\, y \in Y_0} K(y, s) = 0, \qquad s \in S - E_0. \tag{19}$$

Thus for each $s \in S - E_0$ there is a $\delta(s) > 0$ such that

$$|K(y, s)| \leqq 1, \qquad \text{if } y \in Y_0 \text{ and } \|y\| \leqq \delta(s). \tag{20}$$

For each nonzero $y \in Y_0$ choose $\delta_y(s)$ so that $\delta_y(s)/\|y\|$ is rational and $\delta(s)/2 \leqq \delta_y(s) \leqq \delta(s)$. Then from (17) and (20)

$$\left| \frac{\delta_y(s)}{\|y\|} K(y, s) \right| = \left| K\left(\frac{y\delta_y(s)}{\|y\|}, s \right) \right| \leqq 1$$

holds for $s \in S - E_0$ and $y \in Y_0$, so that

$$|K(y, s)| \leqq \frac{2}{\delta(s)} \|y\|, \qquad s \in S - E_0, y \in Y_0. \tag{21}$$

From (17) and (21) we have for $s \in S - E_0$ that

$$|K(y, s) - K(y', s)| \leqq \frac{2}{\delta(s)} \|y - y'\|, \qquad y \in Y_0, y' \in Y_0, \tag{22}$$

which shows that for each such s there is a uniquely defined point x_s in Γ^* with

$$x_s(y) = K(y, s), \qquad y \in Y_0. \tag{23}$$

It may be supposed that the functional x_s over Γ has been extended without increase of norm to the whole of Y and is defined also for $s \in E_0$ by the equation $x_s = \theta_X$. Thus (10) holds. The equalities (14) and (23) show that (2.1.01) is valid for $y \in Y_0$ and the statement (2.1.01) in its entirety follows immediately from this and Lebesgue's convergence theorem since Y_0 is dense in Y' and (18) is true. The conclusion (2.1.02) follows as in the preceding proof since (10), (18), and (23) hold.

Neither conclusion in the above theorem remains true if only (i) and (ii) are assumed. With regard to (2.1.01) it is sufficient to consider the completely additive and absolutely continuous (with respect to α) function defined to the Hilbert space L_2 in Example 9.4 of [31]. This set function is not an integral in the present sense and therefore no function x_s exists satisfying (2.1.01) when $Y' = L_2$. Example 7 of [2] exhibits an integrable function x_s defined to L_2 from $[0, 1]$ for which $\|x_s\|$ is not summable. Here the vector integral $x_E = \int_E x_s d\alpha$ satisfies (i), (ii), and (2.1.01) for $Y' = L_2$ and yet $\|x_s\|$ is not summable.

The following corollary is fundamental for this chapter.

THEOREM 2.1.1. *Let X be the adjoint of another space Y. Suppose x_E is an additive function defined from $\mathcal{E}_B$ to X and that a finite constant K exists such that $\|x_E\| \leqq K\alpha(E)$ for $E \in \mathcal{E}_B$. If Y' is a separable subset of Y, there exists a function x_s defined from S to X and possessing these two properties:*

$$\text{ess. sup. } \|x_s\| \leqq K, \tag{2.1.11}$$

$$x_E(y') = \int_E x_s(y') d\alpha, \qquad E \in \mathcal{E}_B,\ y' \in Y'. \tag{2.1.12}$$

That is, for each such subset Y' in Y there is an essentially bounded function which is Y'-integrable over each $E \in \mathcal{E}_B$ to the value x_E.

A particular case of Theorem 2.1.1 is

THEOREM 2.1.2. *Suppose X is the adjoint Y^* of a separable space Y and that x_E is an additive function defined from $\mathcal{E}_B$ to X and Lipschitzean with constant K. Then there exists a function x_s from S to X such that* (i) ess. sup. $\|x_s\| \leqq K$ *and* (ii) $x_E(y) = \int_E x_s(y) d\alpha$ *for every $y \in Y$ and $E \in \mathcal{E}_B$. Omitting sets of measure zero in S the function x_s is unique.*

That x_s exists satisfying (i) and (ii) is obvious from Theorem 2.1.1. If x_s'

is another such function, pick $\{y_n\}$ dense in Y. From (ii) it follows that $\int_E x_s(y_n)d\alpha = \int_E x_s'(y_n)d\alpha$ or every $E \in \mathcal{E}_B$ and every n, so that $x_s(y_n) = x_s'(y_n)$ holds a.e. in S for each n. This implies that $x_s(y_n) = x_s'(y_n)$, $n = 1, 2, \cdots$, is true for every s in a set S_1 such that $S - S_1 \in \mathcal{E}_0$. The sequence $\{y_n\}$ being dense in Y, the conclusion is that $x_s = x_s'$ in X for each $s \in S_1$, and hence $x_s = x_s'$ a.e. in S.

Theorem 2.1.2 may be restated as follows.

THEOREM 2.1.2′. *From the assumptions in Theorem* 2.1.2 *it follows that there exists an* $x. \in \mathfrak{L}_0^\infty(S)[Y^*, Y]$ *such that* ess. sup. $\|x_s\| \leqq K$ *and* x_s *is* Y*-integrable to* x_E *for each* $E \in \mathcal{E}_B$. *This function* x_s *is essentially unique.*

Suppose the function $x.$ given in Theorem 2.1.2′ is almost separably-valued. Since Y is a determining manifold for $X = Y^*$, Theorem 1.1.9 implies that $x. \in \mathfrak{A}^\infty(S)[X]$. Thus we can state

THEOREM 2.1.3. *If under the assumptions of Theorem* 2.1.2 *the resulting function* $x.$ *is almost separably-valued, then* $x. \in \mathfrak{A}^\infty(S)[X]$ *with* ess. sup. $\|x_s\| \leqq K$ *and* $x_E = \int_E x_s d\alpha$ *for* $E \in \mathcal{E}_B$. *Omitting sets of measure zero* $x.$ *is unique.*

Since Y is separable if its adjoint X is separable, two corollaries result quickly from Theorem 2.1.3.

THEOREM 2.1.4. *If* X *is a separable adjoint space and* x_E *is an additive function from* $\mathcal{E}_B$ *to* X *satisfying a Lipschitz condition with constant* K, *there is an* $x. \in \mathfrak{A}^\infty(S)[X]$ *such that* (i) ess. sup. $\|x_s\| \leqq K$ *and* (ii) $x_E = \int_E x_s d\alpha$, $E \in \mathcal{E}_B$.

THEOREM 2.1.5. *Suppose* X *is a reflexive space and* x_E *fulfills the assumptions of Theorem* 2.1.4. *Then the conclusions of that theorem hold provided that the values of* x_E *form a separable set in* X.

Theorem 2.1.5 is implied by Theorem 2.1.4. For the functional values of x_E lie together in a separable c.l.m. X' in X. Since X is reflexive so is† X', and hence X' is a separable adjoint space.

B. Abstract representations of operations to certain adjoint spaces. The preceding results will now be applied to obtain, in terms of abstract integrals, representations for the general operation sending $L(S)$ into certain adjoint spaces.

THEOREM 2.1.6. *Suppose* X *is the adjoint* Y^* *of a separable space* Y. *If* $x. \in \mathfrak{L}^\infty(S)[Y^*, Y]$, *then* $x. \in \mathfrak{L}_0^\infty(S)[Y^*, Y]$ *and the* Y*-integral*

$$U(\phi) = \int_S x_s \phi(s) d\alpha \tag{2.1.6*}$$

† By a theorem of Plessner; see [16, 18].

defines an operation U from $L(S)$ to X. The function x_s is essentially bounded and $|U| = \text{ess. sup.} \|x_s\|$.

Conversely, if U is an operation from $L(S)$ to X, there is an essentially unique x_s in $\mathfrak{L}_0^\infty(S)[Y^*, Y]$ such that $U(\phi) = \int_S x_s \phi(s) d\alpha$, the integral being the Y-integral. The norm of U is $|U| = \text{ess. sup.} \|x_s\|$.

The first part of this is included in Theorem 1.1.5. To establish the second half let $x_E = U(\phi_E)$ where ϕ_E is the characteristic function of $E \in \mathcal{E}_B$. Then x_E is a set function additive and Lipschitzean over $\mathcal{E}_B$. From Theorem 2.1.2 it follows that there is an essentially unique $x_s \in \mathfrak{L}_0^\infty(S)[Y^*, Y]$ having its Y-integral over E coinciding with x_E for every $E \in \mathcal{E}_B$. But since x_s is in $\mathfrak{L}_0^\infty(S)[Y^*, Y]$, by the first part of the theorem the Y-integral $\int_S x_s \phi(s) d\alpha$ defines a second operation $U'(\phi)$ from $L(S)$ to X, with $|U'| = \text{ess. sup.} \|x_s\|$. Since U and U' are linear and $U(\phi_E) = U'(\phi_E)$ for each $E \in \mathcal{E}_B$, U and U' are identical over $L(S)$. Therefore $U(\phi) = \int_S x_s \phi(s) d\alpha$, where $x_s \in \mathfrak{L}_0^\infty(S)[Y^*, Y]$, and x_s is essentially unique and essentially bounded with ess. sup. $\|x_s\| = |U|$.

Under certain circumstances an operation U from $L(S)$ to X may have a stronger representation, in the sense that the Y-integral in (2.1.6*) becomes an integral. For example there is

THEOREM 2.1.7. *If U is an operation from $L(S)$ to the adjoint X of a separable space Y and if the function x_s defining U by means of* (2.1.6*) *is almost separably-valued, then $x_s \in \mathfrak{A}^\infty(S)[X]$ and $U(\phi) = \int_S x_s \phi(s) d\alpha$. The norm of U is $|U| = \text{ess. sup.} \|x_s\|$.*

From Theorem 1.1.7 and the properties of x_s it is clear that $x_s \in \mathfrak{A}^\infty(S)[X]$ and that the integral $U'(\phi) = \int_S x_s \phi(s) d\alpha$ exists and defines an operation from $L(S)$ to X with $|U'| = \text{ess. sup.} \|x_s\|$. On the other hand, $U(\phi)$ is the Y-integral of $x_s \phi(s)$ for each $\phi(.)$. Since Y's being a determining manifold for $X = Y^*$ implies that this Y-integral coincides with the integral $\int_S x_s \phi(s) d\alpha$, we conclude that $U(\phi) = \int_S x_s \phi(s) d\alpha$ for each ϕ, proving the theorem.

Theorems 2.1.6 and 2.1.7. clearly yield

THEOREM 2.1.8. *Let X be the adjoint Y^* of some space Y and suppose X is separable. If x_s is any element of $\mathfrak{L}^\infty(S)[Y^*, Y]$, then $x_s \in \mathfrak{A}^\infty(S)[X]$, and $U(\phi) = \int_S x_s \phi(s) d\alpha$ is an operation from $L(S)$ to X with $|U| = \text{ess. sup.} \|x_s\|$.*

Conversely, given an operation U from $L(S)$ to this space X there is an essentially unique x_s such that $x_s \in \mathfrak{A}^\infty(S)[X]$ and $U(\phi) = \int_S x_s \phi(s) d\alpha$.

This in turn leads to

THEOREM 2.1.9. *If X has a weakly compact unit sphere, an operation U is defined and separable from $L(S)$ to X if and only if there is an $x_s \in \mathfrak{A}^\infty(S)[X]$*

such that $U(\phi)=\int_S x_s\phi(s)d\alpha$. *The norm of* U *is* $|U|=$ ess. sup. $\|x_s\|$, *and* $x.$ *is essentially unique.*

Let X' be the separable span of $U(L(S))$ in X. Since X' is separable and has a weakly compact unit sphere, it is a separable adjoint space [1, p. 199]. By Theorem 2.1.8 there is an essentially unique $x.\ \varepsilon\ \mathfrak{M}^\infty(S)[X']$ such that $U(\phi)=\int_S x_s\phi(s)d\alpha$. Moreover $|U|=$ ess. sup. $\|x_s\|$. These considerations prove the theorem when combined with the corollary to Theorem 1.2.4.

Part 2. Representations of separable operations to $L^q(T)$, $1<q\leqq\infty$

For Euclidean S and T the general operation from $L(S)$ to $L^q(T)$, $1\leqq q\leqq\infty$, has been represented by B-space-valued functions [18, 9, 23, 24] and by kernel integrals [18, 9, 23, 42, 24, 25]. For $q>1$ the representations by means of vector integrals and kernel integrals due to Gelfand [18] and Dunford [9] are here extended to abstract S and T for which either $\mathcal{E}_B$ or $\mathcal{J}_B$ is separable. The case $q=1$ will be considered in Parts 3 and 4.

A. **Operations with range in** $L^q(T)$, $1<q<\infty$. The space $L^q(T)$ having a weakly compact unit sphere when $1<q<\infty$, Theorem 2.1.9 has as a corollary [18, 9]

Theorem 2.2.1. *An operation* U *is defined and separable from* $L(S)$ *to* $L^q(T)$, $1<q<\infty$, *if and only if there exists an* $x.\ \varepsilon\ \mathfrak{M}^\infty(S)[L^q(T)]$ *such that* $U(\phi)=\int_S x_s\phi(s)d\alpha$. *Here* $x.$ *is essentially unique and* $|U|=$ ess. sup. $\|x_s\|$.

A corresponding theorem in terms of kernels is the following [9].

Theorem 2.2.2. *Given a separable operation* U *from* $L(S)$ *to* $L^q(T)$, $1<q<\infty$, *there is a kernel* $K(s,t)$ *such that*

$$U(\phi)=\int_S K(s,t)\phi(s)d\alpha,\qquad \phi\ \varepsilon\ L(S),\tag{2.2.2*}$$

and

(i) $K(s,t)$ *is measurable, that is,* $\mathcal{E}\times\mathcal{J}$ *measurable, over* $S\times T$,

(ii) ess. sup. $(\int_T|K(s,t)|^q d\beta)^{1/q}\equiv M<\infty$,

(iii) *there exist an* $E_0\ \varepsilon\ \mathcal{E}_0$ *and a separable c.l.m.* $Y\subset L^q(T)$ *such that* $K(s,.)\ \varepsilon\ Y$ *for* $s\notin E_0$.

Conversely, if $K(s,t)$ *satisfies* (i) *and if*

(iv) $K(s,.)\ \varepsilon\ L^q(T)$ *for almost every* s, *and*

(v) ess. sup. $\left|\int_T K(s,t)\psi'(t)d\beta\right|<\infty$ *for each* $\psi'\ \varepsilon\ L^{q'}(T)$,

then $K(s,t)$ *satisfies* (i)–(iii) *and the operation* U *of* (2.2.2*) *is defined and separable from* $L(S)$ *to* $L^q(T)$.

In either case the norm of U *is the constant* M *in* (ii).

Any separable operation U from $L(S)$ to $L^q(T)$, $1<q<\infty$, must have a representation $U(\phi)=\int_S x_s\phi(s)d\alpha$ where $x_\cdot \in \mathfrak{A}^\infty(S)[L^q(T)]$. From Theorem 1.3.5 there exists a kernel $K(s,t)$ satisfying (i), (ii), and (2.2.2*) and such that

$$x_s = K(s,\cdot) \quad \text{in } L^q(T) \text{ for each } s. \tag{1}$$

Since $x_\cdot \in \mathfrak{A}^\infty(S)[L^q(T)]$, (1) implies that $K(s,t)$ also satisfies (iii). From Theorem 1.3.5, $|U|=M$. The second half of the theorem is included in Theorem 1.3.7.

For Euclidean S and T the following particular case of Theorem 2.2.2 has already been established by methods depending on the differentiation either of B-space-valued functions [9] or of real functions [25].

THEOREM 2.2.3. *If either $\mathcal{E}_B$ or $\mathcal{J}_B$ is separable, it follows that U is an operation from $L(S)$ to $L^q(T)$, $1<q<\infty$, if and only if* (2.2.2*) *holds for some kernel satisfying* (i) *and* (ii) *of Theorem* 2.2.2. *The norm of U is $|U|=M$.*

B. **Operations with range in $L^\infty(T)$.** On recalling that $L^\infty(T)$ is the adjoint of $L(T)$ [1, 30, 10] and that $L(T)$ is separable if $\mathcal{J}_B$ is separable, it is obvious that Theorem 2.1.6 yields

THEOREM 2.2.4. *If $\mathcal{J}_B$ is separable, a necessary and sufficient condition that U be an operation defined from $L(S)$ to $L^\infty(T)$ is that there exist an $x_\cdot \in \mathfrak{L}_0^\infty(S)[L^\infty(T), L(T)]$ such that $U(\phi)=\int_S x_s\phi(s)d\alpha$, the integral being the $L(T)$-integral. The function $x_\cdot$ is necessarily essentially bounded and essentially unique, and $|U|=\text{ess. sup.}\|x_s\|$.*

An analogue to Theorem 2.2.4 in terms of kernels is [42, 18, 24]

THEOREM 2.2.5. *Suppose $\mathcal{J}_B$ to be separable. Then U is an operation from $L(S)$ to $L^\infty(T)$ only if there is a $K(s,t)$ having the properties*

(i) *$K(s,t)$ is measurable,*

(ii) *$K(s,t)$ is essentially bounded,*

(iii) *for every ϕ in $L(S)$*

$$U(\phi) = \int_S K(s,t)\phi(s)d\alpha. \tag{2.2.5*}$$

Conversely, if K is a given measurable kernel satisfying the conditions

(iv) $\text{ess. sup.}_t |K(s,t)|<\infty$ *for almost every s, and*

(v) $\text{ess. sup.}_s \left|\int_T K(s,t)\psi'(t)d\beta\right|<\infty$ *for every $\psi' \in L(T)$,*

then K satisfies (ii) *and the mapping U given in* (2.2.5*) *exists and is an operation from $L(S)$ to $L^\infty(T)$.*

The norm of U is $|U|=\text{ess. sup.}|K(s,t)|$.

Consider an arbitrary operation $U(\phi)=\mu_\phi$ from $L(S)$ to $L^\infty(T)$ where $\mathcal{J}_B$

is separable. Let $x.$ be the element of $\mathfrak{L}_0^\infty(S)[L^\infty(T), L(T)]$ associated with U according to Theorem 2.2.4. Taking a decomposition $\{T_j\}$ of T, for each j let x_s^j be the "projection" of x_s onto $L^\infty(T_j)$, that is, x_s^j for each s is the essentially bounded F-measurable function over T_j that coincides with x_s everywhere in T_j. Clearly $x^j. \in \mathfrak{L}_0^\infty(S)[L^\infty(T_j), L(T_j)]$ for each fixed j, so that $x_s^j(\psi')$ is measurable in s whenever x_s^j operates on an element ψ' of $L(T_j)$. Since $\beta(T_j) < \infty$ the function $x^j.$ may be considered as defined to $L(T_j)$ where $L(T_j)$ is separable since $\mathfrak{F}_B$ is separable. So considered $x^j.$ is measurable by Theorem 1.1.7. For $L(T_j)$ is separable, and $x_s^j(\psi')$ is measurable for each $\psi' \in L(T_j)$ and hence for each $\psi' \in L^\infty(T_j)$; thus $x^j.$ is separably-valued and weakly measurable as defined to $L(T_j)$. Theorem 1.3.2 implies, then, that a measurable kernel $K_j(s, t)$ exists over $S \times T_j$ such that $x_s^j = K_j(s, .)$ in $L(T_j)$ for each s. Hence for each j and each s we have $K_j(s, t)$ coinciding a.e. in T_j with x_s. Letting $K(s, t) = K_j(s, t)$ for $s \in S$ and $t \in T_j$, K is evidently measurable and, for each s, $K(s, t)$ coincides a.e. in T with x_s.

Since $|U| = \text{ess. sup.} \|x_s\|$ we now have $|U| = \text{ess. sup.}_s \text{ ess. sup.}_t |K(s, t)|$ and hence by the Fubini theorem $|U| = \text{ess. sup.} |K(s, t)|$, so that (ii) is true. Moreover, since $\mu_\phi = U(\phi)$ is the $L(T)$-integral of $x_s\phi(s)$ over S and $x_s = K(s, .)$ in $L^\infty(T)$, we can write

$$\int_T \mu_\phi(t)\psi'(t)d\beta = \int_S \phi(s)\left\{\int_T K(s, t)\psi'(t)d\beta\right\}d\alpha, \qquad \psi' \in L(T), \phi \in L(S);$$

and again applying the Fubini theorem,

$$\int_T \mu_\phi(t)\psi'(t)d\beta = \int_T \psi'(t)\left\{\int_S \phi(s)K(s, t)d\alpha\right\}d\beta, \qquad \psi' \in L(T), \phi \in L(S).$$

This implies that $\mu_\phi(t) = \int_S \phi(s)K(s, t)d\alpha$ a.e. in T, and hence (iii) is true.

If on the other hand K is a measurable kernel fulfilling (iv) and (v), it is clear that the function $x_s \equiv K(s, .)$ is an element of $\mathfrak{L}^\infty(S)[L^\infty(T), L(T)]$. Hence $x. \in \mathfrak{L}_0^\infty(S)[L^\infty(T), L(T)]$ since $L(T)$ is separable. Applying Theorem 2.2.4, the $L(T)$-integral $U(\phi) = \mu_\phi = \int_S x_s\phi(s)d\alpha$ defines an operation U to $L^\infty(T)$ having $|U| = \text{ess. sup.} \|x_s\|$. Since K is measurable and $|U| = \text{ess. sup.}_s \text{ ess. sup.}_t |K(s, t)|$, it follows from the Fubini theorem that K is essentially bounded over $S \times T$ and that $|U| = \text{ess. sup.}_{(s,t)} |K(s, t)|$. Hence if $\phi \in L(S)$ and $\psi' \in L(T)$ the numerical function $K(s, t)\phi(s)\psi'(t)$ is summable over $S \times T$ and by Fubini's theorem

$$\int_T \psi'(t)\left\{\int_S K(s, t)\phi(s)d\alpha\right\}d\beta = \int_S \phi(s)\left\{\int_T K(s, t)\psi'(t)d\beta\right\}d\alpha.$$

Now from the definition of the $L(T)$-integral we also have

$$\int_T \mu_\phi(t)\psi'(t)d\beta = \int_S \phi(s)\left\{\int_T K(s, t)\psi'(t)d\beta\right\}d\alpha$$

for each $\phi \in L(S)$ and every $\psi' \in L(T)$. Thus for each ϕ the equality

$$\int_T \mu_\phi(t)\psi'(t)d\beta = \int_T \psi'(t)\left\{\int_S K(s, t)\phi(s)d\alpha\right\}d\beta$$

holds for every $\psi' \in L(T)$, which implies that $\mu_\phi = \int_S K(s, t)\phi(s)d\alpha$ in $L^\infty(T)$. This establishes the second half of the theorem, including the fact that $|U| = \text{ess. sup. } |K(s, t)|$.

The representation given in Theorem 2.2.5 is a slight variation on a result stated for bounded Euclidean S and T by Gelfand [18]. A proof of Gelfand's theorem has been given by Kantorovitch and Vulich [24] using linear partially ordered spaces and differentiation theory for real functions. Another representation is due to Vulich [42].

As a corollary to Theorem 2.2.5 there is

THEOREM 2.2.6. *When U is separable from $L(S)$ to $L^\infty(T)$ there is a kernel K defined over $S \times T$ and satisfying conditions* (i)–(iii) *of Theorem* 2.2.5. *The norm of U is* $|U| = \text{ess. sup. } |K(s, t)|$.

Since in $L^\infty(T)$ the span X of $U(L(S))$ is a separable c.l.m., there exists in $\mathcal{F}$ a Borel field $\mathcal{F}'$ such that $\mathcal{F}'_B$ is separable and X is a subset of the set consisting of all elements of $L^\infty(T)$ which are $\mathcal{F}'$-measurable.* Thus X is equivalent to a subset of the adjoint of Y where Y is the B-space of all real functions $\mathcal{F}'$-measurable and β-summable over T. Since U is defined to X and $\mathcal{F}'_B$ is separable, the present theorem follows from the last one.

From Theorem 2.2.6 it is clear that Theorem 2.2.5 holds when $\mathcal{F}_B$ is replaced by $\mathcal{E}_B$.

Theorem 2.2.4 cannot be strengthened to read "there exists an $x. \in \mathfrak{A}^\infty(S)[L^\infty(T)]$ such that $U(\phi) = \int_S x_s\phi(s)d\alpha$." Let $S = T = [0, 1]$ and for each s take x_s to be the characteristic function of the interval $[0, s]$. Evidently $x. \in \mathfrak{L}_0^\infty(S)[L^\infty(T), L(T)]$, so that the $L(T)$-integral of $x_s\phi(s)$ defines an operation U to $L^\infty(T)$. If there were an $x.'$ in $\mathfrak{A}^\infty(S)[L^\infty(T)]$ such that $U(\phi) = \int_S x'_s\phi(s)d\alpha$, Theorem 2.2.4 would imply that $x_s = x'_s$ a.e. in S, so that x_s would be measurable and therefore almost separably-valued. But from the definition of x_s this is clearly false.

* This is a consequence of the following theorem: *a necessary condition that a subset X of $L^q(T)$, $1 \leqq q \leqq \infty$, be separable is that there exist in $\mathcal{F}$ a Borel field $\mathcal{F}'$ having $\mathcal{F}'_B$ separable and such that every element of X is $\mathcal{F}'$-measurable. If $q < \infty$ this condition is also sufficient.* The proof of this is omitted since the essential features can be found in Theorem 3 of [11].

PART 3. THE REPRESENTATIONS OF OPERATIONS TAKING $L(S)$ INTO $C^*(T)$ OR $L(T)$

The remainder of this chapter is devoted to representing those operations sending $L(S)$ into† $C^*(T)$ or $L(T)$ when T is an interval on the real axis (Part 3) and to characterizing some subclasses of these operations (Part 4). In Part 3 the approach in the $L(T)$ case is that used by Vulich [42] and Gelfand [18] and consists first of choosing a space X composed of numerical functions which include the indefinite integrals of the elements of $L(T)$ and then of representing the general operation U from $L(S)$ to X in the form of a kernel integral. Each operation from $L(S)$ to $L(T)$ is therefore obtained by differentiating a kernel integral sending $L(S)$ into X. As with these two authors our choice of X forces T to be a real interval (finite or infinite) with end points c and d, $-\infty \leqq c < d \leqq \infty$, and requires β to be Lebesgue measure. Both, either, or none of the points c and d may be in T. Due to Theorem 2.1.6 we are able however to avoid one restriction present in the representation theorem as proved by Vulich and Gelfand, namely, that S be a bounded real interval. The present proof permits S to be arbitrary.

Let $BV(T)$ be the class of those numerical functions $\nu(t)$ each of which is of bounded variation over T; that is, $BV(T)$ consists of those $\nu(t)$ defined over T for which

$$\operatorname{var}_t \nu(t) \equiv \text{l.u.b.}_{\pi} \sum_{1}^{n} |\nu(t_i) - \nu(t_{i-1})| < \infty$$

holds as π varies over all partitions of the form $\pi = [t_0 < t_1 < \cdots < t_n]$ where $t_i \in T$, $i = 0, 1, \cdots, n$. Each $\nu(.) \in BV(T)$ has the limits $\nu(d-0)$ and $\nu(t+0)$ existing for $c \leqq t < d$; if for a fixed $\nu(.) \in BV(T)$ it is also true that $\nu(c+0) = 0$, $\nu(t+0) = \nu(t)$ for every $t \in T$, and $\nu(d-0) = \nu(d)$ in case $d \in T$, then we say that $\nu(.)$ is in $C^*(T)$. With the norm $\|\nu\| = \rho(\nu) + \operatorname{var}_t \nu(t)$, where $\rho(\nu) = |\nu(c)|$ or $|\nu(c+0)|$ according as $c \in T$ or $c \notin T$, $BV(T)$ is a B-space. In $BV(T)$ the set $C^*(T)$ forms a c.l.m. over which $\|\nu\| = \operatorname{var}_t \nu(t)$, and $C^*(T)$ in turn contains the c.l.m. $AC(T)$ composed of the indefinite Lebesgue integrals of the elements of $L(T)$. Let $C(T)$ denote those numerical functions $\gamma(t)$ which are defined and continuous over T and have the two limits $\gamma(c+0)$ and $\gamma(d-0)$ existing finitely. $C(T)$ is a separable B-space under the norm $\|\gamma\| = \sup_t |\gamma(t)|$. Moreover, for each $\gamma(.) \in C(T)$ and each $\nu(.) \in C^*(T)$ the improper Stieltjes integral

$$S\int_c^d \gamma(t)d\nu(t) = \lim_{c',d'} S\int_{c'}^{d'} \gamma(t)d\nu(t), \qquad c' \to c+0,\ d' \to d-0,$$

† The definition of $C^*(T)$ is given in the next paragraph.

exists and defines over $C(T)$ a linear functional having $\text{var}_t\, \nu(t)$ for its norm. Conversely, given a linear functional $f(\gamma)$ over $C(T)$, exactly one element $\nu(.)$ exists in $C^*(T)$ such that $f(\gamma) = S\int_c^d \gamma(t)d\nu(t)$ for every γ.

Thus the c.l.m. $C^*(T)$ is equivalent to the adjoint of the separable space $C(T)$ and $C^*(T)$ contains the indefinite integrals of the elements of $L(T)$. Taking $C^*(T)$ as the space X mentioned above† we can apply Theorem 2.1.6 to obtain for the generic operation U sending $L(S)$ into $C^*(T)$ a kernel representation $U(\phi) = \int_S K(s, t)\phi(s)d\alpha$ where K has certain detailed properties (Theorem 2.3.1). Since $C^*(T) \supset AC(T)$, this leads immediately to the result of Vulich and Gelfand, namely, that U' is an operation from $L(S)$ to $L(T)$ if and only if

$$U'(\phi) = \frac{d}{dt}\int_S K(s, t)\phi(s)d\alpha$$

where $K(s, t)$ belongs to a specified class of kernels (Theorem 2.3.9).

We thus suppose T to be a fixed linear interval and $BV(T)$, $C^*(T)$, and $C(T)$ to be the spaces described above. The set of operations mapping $L(S)$ into $C^*(T)$ will be denoted by $\mathfrak{U}$. Finally, it is to be noted that for each $t \in T$ the functional $f_t(\nu) \equiv \nu(t)$ is linear over $C^*(T)$ and $\|f_t\| = 1$.

A. **Operations to $C^*(T)$.** The chief theorem of this part is

THEOREM 2.3.1. *Any operation $U(\phi) = x_\phi$ from $L(S)$ to $C^*(T)$, that is, any $U \in \mathfrak{U}$, has a representation of the form*

$$U(\phi) = x_\phi = \int_S K(s, t)\phi(s)d\alpha, \qquad \phi \in L(S), \tag{2.3.1*}$$

where $K(s, t)$ is a real kernel defined over $S \times T$ and having the following properties:

(2.3.11) *$K(s, .) \in C^*(T)$ for each $s \in S$,*

(2.3.12) *$K(., t) \in L^\infty(S)$ for each $t \in T$,*

(2.3.13) *$K(s, t)$ is measurable and* ess. sup. $|K(s, t)| < \infty$,

(2.3.14) ess. sup.$_s$ $[\text{var}_t K(s, t)] \equiv M < \infty$,

(2.3.15) *the two iterated integrals $S\int_c^d \gamma(t)d\{\int_S K(s, t)\phi(s)d\alpha\}$ and $\int_S \{S\int_c^d \gamma(t)d_t K(s, t)\}\phi(s)d\alpha$ exist and are equal whenever $\gamma \in C(T)$ and $\phi \in L(S)$.*

Conversely, if $K(s, t)$ satisfies the conditions

(2.3.16) *$K(s, .) \in C^*(T)$ for almost every s,*

(2.3.17) *$K(s, t)$ is measurable in s for every t in a dense subset of T,*

(2.3.18) ess. sup.$_s$ $\left|S\int_c^d \gamma(t)d_t K(s, t)\right| < \infty$ *for each $\gamma(.) \in C(T)$,*

† In [41] and [18] $BV(T)$ is chosen as the space X. For a kernel representation of the general operation from $L(S)$ to $BV(T)$ see Kantorovitch [23], Vulich [42], and Gelfand [18].

then K has the additional properties (2.3.12)–(2.3.15) *and the operation U of* (2.3.1*) *is defined and linear from $L(S)$ to $C^*(T)$, that is, $U \in \mathfrak{U}$.*

The norm of U in either case is the constant M of (2.3.14).

The proof will consist of a sequence of four theorems each of which will also be useful in later pages. The first of these is an obvious consequence of Theorem 2.1.6.

THEOREM 2.3.2. *A mapping U is an operation $U(\phi) = \nu_\phi$ defined from $L(S)$ to $C^*(T)$ if and only if there is a $\nu. \in \mathfrak{L}^\infty(S)[C^*(T), C(T)] = \mathfrak{L}_0^\infty(S)[C^*(T), C(T)]$ such that, in $C^*(T)$, ν_ϕ is the $C(T)$-integral $\int_S \nu_s \phi(s) d\alpha$ for each $\phi \in L(S)$, that is,*†

$$(2.3.2^*) \qquad \nu_\phi \cdot \gamma = \int_S \nu_s \cdot \gamma \phi_s d\alpha, \qquad \gamma \in C(T),\ \phi \in L(S).$$

This function $\nu.$ is essentially unique and ess. sup. $\|\nu_s\| = |U|$.

The second theorem is important.

THEOREM 2.3.3. *Suppose that $K(s, t)$ is a numerical kernel given over $S \times T$ and that there exist an $E_0 \in \mathcal{E}_0$ and a $\nu. \in \mathfrak{L}_0^\infty(S)[C^*(T), C(T)]$ such that*

$$(2.3.31) \qquad K(s, .) = \nu_s \text{ in } C^*(T) \qquad \text{for each } s \in S - E_0.$$

Let U be the element of $\mathfrak{U}$ defined by the $C(T)$-integral $U(\phi) = \nu_\phi = \int_S \nu_s \phi(s) d\alpha$. Then K has the properties (2.3.12)–(2.3.16) *and $U(\phi) = \int_S K(s, t)\phi(s) d\alpha$ for each $\phi \in L(S)$. The norm of U is the constant M of* (2.3.14).

In the proof of this the following lemma will be needed.

LEMMA. *If $t' \in T$ is fixed and $f_{t'}(\nu)$ is the linear functional over $C^*(T)$ defined by $f_{t'}(\nu) \equiv \nu(t')$, then there exists in $C(T)$ a sequence $\{\gamma_n\}$ such that $\|\gamma_n\| \leqq 1$ and $f_{t'}(\nu) = \lim_n \mathrm{S}\int_c^d \gamma_n(t) d\nu(t)$ for every $\nu \in C^*(T)$. The sequence $\{\gamma_n\}$ is independent of ν and depends only on t'.*

If $t' = c$ or $t' = d$ the respective sequences $\gamma_n(t) \equiv 0$ or $\gamma_n(t) \equiv 1$ will serve. In case $c < t' < d$ let $\{h_n\}$ be positive constants such that $T \supset \{t' + h_n\}$ and $\lim h_n = 0$. For each n define $\gamma_n(t)$ to have the value 1 for $c < (=)\ t \leqq t'$, the value 0 for $t' + h_n \leqq t < (=)\ d$, and to be linear between t' and $t' + h_n$. Clearly $\|\gamma_n\| \leqq 1$ and $\mathrm{S}\int_c^d \gamma_n(t) d\nu(t) = \mathrm{S}\int_c^d \gamma_n(t) df_t(\nu) = f_{t'}(\nu) + \mathrm{S}\int_{t'}^{t'+h_n} \gamma_n(t) df_t(\nu)$ for arbitrary $\nu \in C^*(T)$; thus it is sufficient to prove that $0 = \lim_n \mathrm{S}\int_{t'}^{t'+h_n} \gamma_n(t) df_t(\nu)$. There is no loss of generality in supposing the element $\nu(.)$ of $C^*(T)$ to be monotone non-decreasing as a function of t. For a fixed ν of this sort we can then write, due to h_n's being positive,

† For each $\nu \in C^*(T)$ and each $\gamma \in C(T)$ the symbol $\nu \cdot \gamma$ represents the value taken at γ by the linear functional defined over $C(T)$ by ν, that is, $\nu \cdot \gamma$ is the value of $\mathrm{S}\int_c^d \gamma(t) d\nu(t)$.

$$\left| S\int_{t'}^{t'+h_n} \gamma_n(t)df_t(\nu) \right| \leqq \max_t |\gamma_n(t)| \cdot \operatorname{var}_t [f_t(\nu); t', t'+h_n]$$
$$= f_{t'+h_n}(\nu) - f_{t'}(\nu),$$

and hence $\lim_n \int_{t'}^{t'+h_n} \gamma_n(t)df_t(\nu) = 0$, since $\nu(t') = \nu(t'+0)$. This establishes the lemma.

In Theorem 2.3.3 it is obvious, in view of (2.3.31), that $K(s, t)$ has property (2.3.16). Moreover, (2.3.31) and Theorem 2.3.2 imply that the $C(T)$-integral $\int_S \nu_s \phi(s)d\alpha = \nu_\phi$ defines an operation $U(\phi) = \nu_\phi$ from $L(S)$ to $C^*(T)$ with $|U| = \text{ess. sup.} \|\nu_s\| = \text{ess. sup.}_s [\text{var}_t K(s, t)]$; thus $K(s, t)$ satisfies (2.3.14). It is also evident from (2.3.31) and (2.3.2*) that

$$(1) \qquad S\int_c^d \gamma(t)d_t f_t(\nu_\phi) = \nu_\phi \cdot \gamma = \int_S \nu_s \cdot \gamma \phi(s)d\alpha$$
$$= \int_S \left\{ S\int_c^d \gamma(t)d_t K(s, t) \right\} \phi(s)d\alpha$$

holds for every $\gamma \in C(T)$ and every $\phi \in L(S)$. To establish (2.3.12) fix $t' \in T$. According to the lemma there is in $C(T)$ a sequence $\{\gamma_n\}$ independent of s and such that $f_{t'}(\nu_s) = \lim_n \nu_s \cdot \gamma_n$ whenever $\nu_s \in C^*(T)$. Since (2.3.31) implies that for almost every s we have both $\nu_s \in C^*(T)$ and $K(s, t') = f_{t'}(\nu_s)$, the conclusion is that $\lim_n \nu_s \cdot \gamma_n = K(s, t')$ a.e. in S. Hence $K(s, t')$ is measurable, since by assumption $\nu_s \cdot \gamma_n$ is measurable for each n. Moreover,

$$(2) \qquad \text{ess. sup.}_s |K(s, t')| \leqq M, \qquad t' \in T,$$

because ess. sup. $|\nu_s \cdot \gamma_n| \leqq \|\gamma_n\|$ ess. sup. $\|\nu_s\| \leqq M$. Thus (2.3.12) is true.

To see that $K(s, t)$ is measurable over $S \times T$, for each m let $t_{m,j}$, $j = \cdots, -1, 0, 1, \cdots$, be points in T such that $0 < t_{m,j} - t_{m,j-1} < 1/2^m$ and $\lim_{j \to -\infty} t_{m,j} = c$, $\lim_{j \to \infty} t_{m,j} = d$. Define $K_m(s, t) = K(s, t_{m,j})$ when $t_{m,j-1} \leqq t < t_{m,j}$. From (2.3.12), $K(s, t_{m,j})$ is measurable in s, so that $K_m(s, t)$ is measurable in $S \times (c, d)$. Since $K(s, t+0) = K(s, t)$ for $t < d$ and $s \in S - E_0$, it follows that $K(s, t) = \lim_m K_m(s, t)$ for each $s \in S - E_0$ and each t not of the form $t = t_{k,j}$, c, or d. Thus $K(s, t)$ is a.e. in S the limit of a sequence of measurable functions and is therefore measurable. The inequality ess. sup. $|K(s, t)| \leqq M$ now results from (2) and the Fubini theorem. This vindicates (2.3.13).

In showing for each ϕ that $U(\phi) = \nu_\phi = \int_S K(s, t)\phi(s)d\alpha$, that is, that $f_t(\nu_\phi) = \int_S K(s, t)\phi(s)d\alpha$ for all t, use is again made of the lemma. Fixing t, sequence $\{\gamma_n\}$ exists in $C(T)$ independently of ϕ and such that $f_t(\nu_\phi) = \lim_n \nu_\phi \cdot \gamma_n$, where $\|\gamma_n\| \leqq 1$. Applying (1) this becomes

$$f_t(\nu_\phi) = \lim_n \int_S \nu_s \cdot \gamma_n \phi(s) d\alpha, \qquad \phi \in L(S). \tag{3}$$

On the other hand the lemma also yields the information that $f_t(\nu_s)\phi(s) = \lim_n \nu_s \cdot \gamma_n \phi(s)$ for almost all s. Since $f_t(\nu_s)\phi(s)$ equals $K(s, t)\phi(s)$ for all $s \in S - E_0$ by (2.3.31), it follows that $K(s, t)\phi(s) = \lim_n \nu_s \cdot \gamma_n \phi(s)$ a.e. in S. From the inequality $|\nu_s \cdot \gamma_n \phi(s)| \leqq M\|\gamma_n\| |\phi(s)| \leqq M|\phi(s)|$ and Lebesgue's convergence theorem, $K(s, t)\phi(s)$ is summable in s and $\int_S K(s, t)\phi(s)d\alpha = \lim_n \int_S \nu_s \cdot \gamma_n \phi(s) d\alpha$. From (3) we can now write

$$f_t(\nu_\phi) = \int_S K(s, t)\phi(s)d\alpha, \qquad t \in T, \phi \in L(S), \tag{4}$$

so that $U(\phi) = \int_S K(s, t)\phi(s)d\alpha$ where $|U| = \text{ess. sup. } \|x_s\| = M$. The proof is complete on noting that property (2.3.15) results immediately from (1) and (4).

THEOREM 2.3.4. *Given $\nu_\cdot \in \mathfrak{L}^\infty(S)[C^*(T), C(T)]$, a kernel $K(s, t)$ exists over $S \times T$ such that* (2.3.11) *is satisfied and* (2.3.31) *is true for some $E_0 \in \mathcal{E}_0$. Thus all the conclusions of Theorem* 2.3.3 *hold for the kernel $K(s, t)$ and the function $\nu_\cdot$.*

Set $E_0 = S[\nu_s \notin C^*(T)]$ and let $K(s, t) = f_t(\nu_s)$ for $s \in S - E_0$ and $t \in T$, $K(s, t) \equiv 0$ for $s \in E_0$. Obviously (2.3.11) is fulfilled, and since $E_0 \in \mathcal{E}_0$ and $\mathfrak{L}^\infty(S)[C^*(T), C(T)] = \mathfrak{L}_0^\infty(S)[C^*(T), C(T)]$ it follows that (2.3.31) holds.

The last of the four theorems needed is

THEOREM 2.3.5. *Let $K(s, t)$ over $S \times T$ satisfy the conditions* (2.3.17) *and* (2.3.18), *and let $E_0 \in \mathcal{E}_0$ be such that $K(s, .) \in C^*(T)$ for $s \in S - E_0$. Then the function $\nu_s \equiv K(s, .)$, $s \in S - E_0$, is in $\mathfrak{L}_0^\infty(S)[C^*(T), C(T)]$ and all the conclusions of Theorem* 2.3.3 *are true.*

Since $\mathfrak{L}_0^\infty(S)[C^*(T), C(T)] = \mathfrak{L}^\infty(S)[C^*(T), C(T)]$, it is sufficient to show that $\nu_s \cdot \gamma \in L^\infty(S)$ for each $\gamma \in C(T)$. For each $s \in S - E_0$, $E_0 \in \mathcal{E}_0$, we have $\nu_s \cdot \gamma = \text{S}\int_c^d \gamma(t) dK(s, t)$, so that ess. sup. $|\nu_s \cdot \gamma| < \infty$ from (2.3.18). And from (2.3.17) it follows that $\nu_s \cdot \gamma$ is measurable in s. For points $\{t_{n,i}\}$, $n = 1, 2, \cdots$, $i = \cdots, -1, 0, 1, \cdots$, can be so chosen in T that $0 < t_{n,i} - t_{n,i-1} < 1/2^n$, $\lim_{i \to -\infty} t_{n,i} = c$, $\lim_{i \to \infty} t_{n,i} = d$, and $K(s, t_{n,i})$ is measurable in s. Setting

$$p_n(\gamma, s) = \sum_{i=-\infty}^{\infty} \gamma(t_{n,i})[K(s, t_{n,i}) - K(s, t_{n,i-1})]$$

it is clear for each $\gamma \in C(T)$ and each n that $p_n(\gamma, s)$ exists for $s \in S - E_0$ and is measurable. Moreover,

$$\nu_s \cdot \gamma = \text{S}\int_c^d \gamma(t) dK(s, t) = \lim_n p_n(\gamma, s)$$

for $s \in S-E_0$. Hence for each γ, $\nu_s \cdot \gamma$ is the limit a.e. of measurable functions, and so is itself measurable. This ends the proof.

The demonstration of Theorem 2.3.1 can now be easily constructed. If U is in $\mathfrak{U}$, there is, by Theorem 2.3.2, a ν. $\in \mathfrak{L}_0^\infty(S)[C^*(T), C(T)]$ such that $U(\phi)$ is the $C(T)$-integral $\int_S \nu_s \phi(s) d\alpha$ for each ϕ. From Theorem 2.3.4 a kernel $K(s, t)$ exists over $S \times T$ having the properties (2.3.11)–(2.3.15) and such that $U(\phi) = \int_S K(s, t)\phi(s) d\alpha$, where $|U|$ is the constant M of (2.3.14). If, on the other hand, $K(s, t)$ is given satisfying (2.3.16)–(2.3.18), then by Theorem 2.3.5 $U(\phi) = \int_S K(s, t)\phi(s) d\alpha$ is an operation from $L(S)$ to $C^*(T)$ with $|U| = M$.

For a given kernel $H(s, t)$ varying sets of conditions can be proved to be sufficient in order that the operation $U(\phi) = \int_S H(s, t)\phi(s) d\alpha$ should be a point in $\mathfrak{U}$. The following substitution theorem may therefore be of some use in that it asserts that each measurable kernel of this sort is equivalent, as far as measure is concerned, to a second kernel which satisfies (2.3.11)–(2.3.15). Moreover this second kernel defines the same operation, so that it can serve as a replacement for $H(s, t)$.

THEOREM 2.3.6. *Given a measurable kernel $H(s, t)$ with the property that $U(\phi) = \int_S H(s, t)\phi(s) d\alpha$ is in $\mathfrak{U}$, there exists a second kernel $K(s, t)$ such that*

(2.3.61) $U(\phi) = \int_S K(s, t)\phi(s) d\alpha, \phi \in L(S)$,

(2.3.62) *$K(s, t)$ satisfies* (2.3.11)–(2.3.15),

(2.3.63) *the function $\nu_s \equiv K(s, .)$ is in $\mathfrak{L}^\infty(S)[C^*(T), C(T)]$*,

(2.3.64) *$t \in T$ implies $K(s, t) = H(s, t)$ a.e. in S*,

(2.3.65) *$K(s, t) = H(s, t)$ a.e. in $S \times T$*,

(2.3.66) *for almost all s the equality $K(s, t) = H(s, t)$ holds a.e. in T.*

According to Theorem 2.3.1 there is a $K(s, t)$ such that (2.3.61) and (2.3.62) are true. Hence for each t the two integrals $\int_S H(s, t)\phi(s) d\alpha$ and $\int_S K(s, t)\phi(s) d\alpha$ coincide for every $\phi \in L(S)$, which implies that, for each t, $K(s, t) = H(s, t)$ a.e. in S; thus (2.3.64) holds. (2.3.65) now follows from the Fubini theorem and the measurability of the two kernels, and (2.3.66) is a consequence of (2.3.65). Finally, since (2.3.11)–(2.3.15) imply (2.3.16)–(2.3.18) it is seen that (2.3.63) follows from (2.3.62) and Theorem 2.3.5.

A given kernel $H(s, t)$ may define a linear operation $\int_S H(s, t)\phi(s) d\alpha$ from $L(S)$ to $C^*(T)$ and yet possess very few desirable properties beyond the obvious ones that $H(., t)$ is in $L^\infty(S)$ for each t and that $K(s, t)$ exists satisfying (2.3.11)–(2.3.15) and having $K(s, t) = H(s, t)$ a.e. in S for each t. As an example, in the unit square let $H(s, t)$ be the characteristic function of Sierpiński's non-measurable set that has at most two points in common with any straight line [36]. Here $\int_S H(s, t)\phi(s) d\alpha$ sends each ϕ into the identically zero function; yet $H(s, t)$ is non-measurable, $|U| \neq$ ess. sup.$_s$ [var$_t$ $H(s, t)$], and

(2.3.16) fails to be true. The final two theorems of this part show that with an additional assumption it can be concluded that $H(s, t)$ does have properties (2.3.12)–(2.3.16).

THEOREM 2.3.7. *Suppose that $H(s, t)$ is measurable, that the mapping $U(\phi) = \int_S H(s, t)\phi(s)d\alpha$ is an element of $\mathfrak{U}$, and that $H(s, t)$ is for almost every s everywhere continuous on the right in t, Then $H(s, t)$ satisfies* (2.3.12)–(2.3.16) *and $\nu_s \equiv H(s, .)$ is in $\mathfrak{L}_0^\infty(S)[C^*(T), C(T)]$.*

If $K(s, t)$ is the kernel given in the conclusion of Theorem 2.3.6 , the equality $H(s, t) = K(s, t)$ holds a.e. in T for almost every s. Since for almost every s it is also true that $H(s, t)$ is continuous on the right over T, it is evident there exists an $E_0 \in \mathcal{E}_0$ such that $s \in S - E_0$ implies $H(s, t) = K(s, t)$ everywhere in T; for two functions of t both continuous on the right and coinciding over a dense set are necessarily identical. Thus $H(s, .)$ is, for almost every s, the element $K(s, .)$ of $C^*(T)$. Since $K(s, .) \in \mathfrak{L}_0^\infty(S)[C^*(T), C(T)]$ by Theorem 2.3.5, it follows from Theorem 2.3.3 that $H(s, t)$ satisfies (2.3.12)–(2.3.16). From this fact and from Theorem 2.3.5 we conclude also that $\nu_s \equiv H(s, .)$ is in $\mathfrak{L}_0^\infty(S)[C^*(T), C(T)]$.

THEOREM 2.3.8. *Suppose $H(s, .)$ is essentially defined to $C^*(T)$ and that $U(\phi) = \int_S H(s, t)\phi(s)d\alpha$ is in $\mathfrak{U}$. Then $\nu_s \equiv H(s, .)$ is in*

$$\mathfrak{L}_0^\infty(S)[C^*(T), C(T)]$$

and all the conclusions of Theorem 2.3.3 hold. In particular $H(s, t)$ satisfies (2.3.12)–(2.3.16) *and hence is measurable.*

It is evident that $H(s, t)$ satisfies (2.3.16) and (2.3.17). In view of Theorem 2.3.5 there is need only of showing that ess. sup.$_s$ $|\mathrm{S}\int_c^d \gamma(t)dK(s, t)| < \infty$ for each $\gamma \in C(T)$. This may be proved as follows. The function $h_t = H(., t)$ is defined from T to $L^\infty(S)$ and $h_t \cdot \phi = \int_S H(s, t)\phi(s)d\alpha$ is of bounded variation in t for each $\phi \in L(S)$. Hence from [10] it is seen that h_t is an abstract function having bounded variation in the sense of that paper, and that the abstract integral $V(\gamma) = \mathrm{S}\int_c^d \gamma(t) \cdot d_t h_t$ exists as an element of $L^\infty(S)$ for each $\gamma \in C(T)$. Fix γ and in T choose points $\{t_{n,i}\}$, $n = 1, 2, \cdots, i = \cdots, -1, 0, 1, \cdots$, satisfying the conditions $0 < t_{n,i} - t_{n,i-1} < 1/2^n$ and $\lim_{i \to -\infty} t_{n,i} = c$, $\lim_{i \to \infty} t_{n,i} = d$. For each s such that $H(s, .) \in C^*(T)$ the function $\mu_n(s) = \sum_i \gamma(t_{n,i})[H(s, t_{n,i}) - H(s, t_{n,i-1})]$ is defined and $\lim_n \mu_n(s) = \mathrm{S}\int_c^d \gamma(t)d_t H(s, t)$. On the other hand from the definition of $\mathrm{S}\int_c^d \gamma(t)d_t h_t$ we have $V(\gamma) = \lim_n \mu_n(.)$ in $L^\infty(S)$, so that $\{\mu_n(s)\}$ converges a.e. in S to $\mu(s)$ where $\mu(.) = V(\gamma)$. Since $H(s, .) \in C^*(T)$ for almost all s, we now have $\mathrm{S}\int_c^d \gamma(t)dH(s, t) = \mu(s)$ holding a.e. in S, and hence ess. sup. $|\mathrm{S}\int_c^d \gamma(t)dH(s, t)| < \infty$. This establishes Theorem 2.3.8.

As an immediate corollary we see that if $H(s, .)$ and $K(s, .)$ are essentially defined to $C^*(T)$ and define the same operation $U \in \mathfrak{U}$ then $H(s, t) = K(s, t)$, $t \in T$, $s \in S - E_0$ for some $E_0 \in \mathcal{E}_0$.

B. **Operations from $L(S)$ to $L(T)$.** From Theorem 2.3.1 the reader can easily infer [42, 18]

THEOREM 2.3.9. *Any operation U' from $L(S)$ to $L(T)$ has a representation of the form*

$$U'(\phi) = \psi_\phi = \frac{d}{dt}\int_S K(s, t)\phi(s)d\alpha, \qquad \phi \in L(S), \tag{2.3.9*}$$

where $\int_S K(s, t)\phi(s)d\alpha \in AC(T)$ for each ϕ and $K(s, t)$ has properties (2.3.11)–(2.3.15). *The norm of U' is the constant M of* (2.3.14).

Conversely, if $K(s, t)$ satisfies (2.3.16)–(2.3.18), *the operation U' of* (2.3.9*) *is defined from $L(S)$ to $L(T)$ and $K(s, t)$ has properties* (2.3.12)–(2.3.15). *Here $|U'| \leqq M$, with equality holding if $\int_E K(s, t)d\alpha \in AC(T)$ for each $E \in \mathcal{E}_B$.*

PART 4. SUBCLASSES OF THOSE OPERATIONS HAVING THEIR RANGE IN $C^*(T)$ OR $L(T)$

In the preceding it was seen that any operation U' from $L(S)$ to $L(T)$ has a representation

$$U'(\phi) = \frac{d}{dt}\int_S H(s, t)\phi(s)d\alpha \tag{2.4.01}$$

where H is measurable and $H(s, .)$ is defined to $BV(T)$ (to $C^*(T)$ actually). Since some operations mapping $L(S)$ into $L(T)$ cannot have the form

$$U'(\phi) = \int K'(s, t)\phi(s)d\alpha \tag{2.4.02}$$

where K' is measurable, the question naturally arises as to when a given U' does have the more convenient representation (2.4.02). In this part an answer is given in the following terms. U' can be written in the form (2.4.02) with K' measurable and ess. $\sup._s \int_T |K'(s, t)|\, d\beta < \infty$ if and only if among the representations (2.4.01) of U' there is one with H measurable and $H(s, .)$ essentially defined to some separable c.l.m. in $BV(T)$. This result is obtained by establishing in terms of kernels a necessary and sufficient condition that a given $U \in \mathfrak{U}$ be representable as $U(\phi) = \int_S \nu_s\phi(s)d\alpha$ where $\nu_. \in \mathfrak{A}^\infty(S)[C^*(T)]$. In the third section characterizations by means of kernels are given for the subclass in $\mathfrak{U}$ composed of those operators defined by elements of $\mathfrak{A}^\infty(S)[AC(T)]$.

A. **Operations to $C^*(T)$ given by almost separably-valued kernels.** We begin with

THEOREM 2.4.1. *Suppose $U \in \mathfrak{U}$ has a representation $U(\phi) = \int_S \nu_s \phi(s) d\alpha$ where $\nu_\cdot \in \mathfrak{A}^\infty(S)[C^*(T)]$. In $C^*(T)$ let Y be the separable span of $U(L(S))$. Then*

(i) *$\nu_\cdot$ is essentially defined to Y,*

(ii) *U can be written as*

$$U(\phi) = \int_S K(s, t)\phi(s) d\alpha$$

where $K(s, t)$ has properties (2.3.11)–(2.3.15) and $K(s, .) = \nu_s$ for almost every s,

(iii) *$K(s, .)$ is essentially defined to Y and $K(s, .)$ is almost separably-valued in $C^*(T)$. The norm of U is $|U| = \text{ess. sup.} \|\nu_s\| = \text{ess. sup.}_s [\text{var}_t K(s, t)]$.*

From the corollary to Theorem 1.2.4 the span Y is separable in $C^*(T)$ and by Theorem 1.2.9 $\nu_\cdot$ is essentially defined to Y. The remaining conclusions can now be derived from Theorem 2.3.4.

In the direction reverse to Theorem 2.4.1 there is the more difficult

THEOREM 2.4.2. *Let $U(\phi) = \int_S H(s, t)\phi(s) d\alpha$ be an operation from $L(S)$ to $C^*(T)$ where H is measurable and $H(s, .)$ is essentially defined to a separable c.l.m. X in $BV(T)$. Then*

(i) *$H(s, .)$ is essentially defined to the separable c.l.m. $Y = C^*(T) \cdot X$ in $C^*(T)$,*

(ii) *$H(s, t)$ has properties (2.3.12)–(2.3.16),*

(iii) *U has a representation $U(\phi) = \int_S \nu_s \phi(s) d\alpha$ where $\nu_\cdot \in \mathfrak{A}^\infty(S)[Y]$ and $H(s, .) = \nu_s$ in $C^*(T)$ for almost every s, and*

(iv) *the range of U is in Y and $|U| = \text{ess. sup.} \|\nu_s\| = \text{ess. sup.}_s [\text{var}_t H(s,t)]$.*

Let $K(s, t)$ be a kernel representing U according to Theorem 2.3.6 so that for almost every s we have $K(s, t) = H(s, t)$ a.e. in T. Since $H(s, .)$ is essentially defined to X, there is then a null set E_0' such that for each $s \in S - E_0'$

$$H(s, t) = K(s, t) \text{ a.e. in } T, \tag{1}$$

$$H(s, .) \in X \subset BV(T). \tag{2}$$

The subset X being separable, a denumerable set $\{s_n\}$ exists in $S - E_0'$ with the property that for each $s \notin E_0'$ there is a subsequence $\{s_{m_i}\}$ such that $\lim_{i\to\infty} \|H(s, .) - H(s_{m_i}, .)\| = 0$, the norm being taken in $BV(T)$. This implies that

$$\lim_{i\to\infty} H(s_{m_i}, t) = H(s, t) \qquad \text{uniformly in } t. \tag{3}$$

Since $H(s_m, t)$ is of bounded variation in t, the set of points in T for which the equality $H(s_m, t) = H(s_m, t+0)$ holds for every m includes all of T except an at most denumerable set. Let $\{t_n\}$ be this exceptional set plus any end points of T that may be in T. From (2.3.64) there is for each n a null set E_n such that $H(s, t_n) = K(s, t_n)$ for $s \notin E_n$. Let $E_0 = E_0' + \sum E_n$, so that E_0 is null. For each $s \in S - E_0$ we now have not only (1)–(3) holding but also

$$H(s, t_n) = K(s, t_n), \qquad n = 1, 2, \cdots. \tag{4}$$

From (3) it is seen that if $s \notin E_0$ and $H(s, t^*+0) \neq H(s, t^*)$ for some t^* then $H(s_{m_i}, t^*+0) \neq H(s_{m_i}, t^*)$ for at least one i. This of course means that $t^* = t_n$ for at least one n. Hence we can conclude that for $s \notin E_0$

$$H(s, t+0) = H(s, t), \qquad t \notin \{t_n\}. \tag{5}$$

For each fixed $s \in S - E_0$ it now follows that $H(s, t) = K(s, t)$ for all t. For (1) implies that $H(s, t) = K(s, t)$ over a dense set in T so that

$$H(s, t+0) = K(s, t+0) \qquad \text{for every } t. \tag{6}$$

From (4), (5), (6) and the right continuity of $K(s, t)$ everywhere in T it is then seen that

$$K(s, t) = K(s, t+0) = H(s, t+0) = H(s, t)$$

holds for all t. Since $s \in S - E_0$ implies $K(s, .) \in C^*(T)$ and $H(s, .) \in X$, $H(s, .)$ is essentially defined to Y. This establishes (i). Conclusion (ii) follows from (i) and Theorem 2.3.8. To obtain (iii) we set $\nu_s = K(s, .)$ and note first that $\nu_s = H(s, .)$ for almost all s. From Theorem 2.3.5 and the equality $\nu_s = K(s, .)$ it follows that $\nu_. \in \mathfrak{L}_0^\infty(S)[C^*(T), C(T)]$. Since ν_s coincides a.e. over S with the almost separably-valued function $H(s, .)$, Theorem 1.1.9 then implies that $\nu_. \in \mathfrak{L}^\infty(S)[C^*(T)]$ and Theorem 2.3.3 asserts that $\int_S H(s, t)\phi(s)d\alpha = \int_S \nu_s\phi(s)d\alpha$ in $C^*(T)$. Thus (iii) is proved. It is a result of (iii) and Theorem 1.2.4 that the range of U is in Y. The theorem is established.

With each $U \in \mathfrak{U}$ there are associated two other operations, $U'(\phi) = d\nu_\phi(t)/dt$ with range in $L(T)$ and

$$V(\phi) = \int_c^\tau \left[\frac{d}{dt}\nu_\phi(t)\right] dt$$

with range in $AC(T)$. The first of these can always be represented in the form (2.3.9*) but not always can it be written as $U'(\phi) = \int_S K'(s, t)\phi(s)d\alpha$ where K' is measurable. In the next theorem a condition is given which is sufficient that such a $K'(s, t)$ exist.

THEOREM 2.4.3. *From the assumptions of Theorem* 2.4.2 *it follows that the associated operation*

$$U'(\phi) = \frac{d}{dt}\int_S H(s, t)\phi(s)d\alpha$$

from $L(S)$ *to* $L(T)$ *can be represented as*

$$U'(\phi) = \int_S \psi_s\phi(s)d\alpha \tag{2.4.31}$$

where $\psi_\cdot \in \mathfrak{A}^\infty(S)[L(T)]$, *and as*

$$U'(\phi) = \int_S K'(s, t)\phi(s)d\alpha \tag{2.4.32}$$

where (i) $K'(s, t)$ *is measurable,* (ii) $K'(s, .) = \psi_s$ *is in* $L(T)$ *for every* s, *and* (iii) *for almost every* s, $K'(s, t) = dH(s, t)/dt$ *a.e. in* T. *The norm of* U' *is* $|U'| = \text{ess. sup.}_s \int_T |K'(s, t)|\,d\beta$.

Let $W(\nu) = \psi$ assign to each $\nu(.) \in C^*(T)$ its derivative function $\psi(t) = d\nu(t)/dt$. Then $U'(\phi) = W(U(\phi)) = W(\int_S \nu_s\phi(s)d\alpha)$ where $\nu_\cdot \in \mathfrak{A}^\infty(S)[C^*(T)]$ and $\nu_s = H(s, .)$ a.e. in S. Hence [2] $U'(\phi) = \int_S W(\nu_s)\phi(s)d\alpha$, and $\psi_s \equiv W(\nu_s)$ is in $\mathfrak{A}^\infty(S)[L(T)]$. Thus $U'(\phi) = \int_S \psi_s\phi(s)d\alpha$ and $|U'| = \text{ess. sup.} \|\psi_s\|$. Theorem 1.3.5 now provides a measurable kernel $K'(s, t)$ such that $K'(s, .)$ is in $L(T)$ for each s, $U'(\phi) = \int_S K'(s, t)\phi(s)d\alpha$ for every $\phi \in L(S)$, and $|U'| = \text{ess. sup.}_s \int_T |K'(s, t)|\,d\beta$. Finally, for almost every s we have $W(H(s, .)) = W(\nu_s) = \psi_s = K'(s, .)$ and so $dH(s, t)/dt = K'(s, t)$ a.e. in T.

The preceding results may be collected in the following statement.

THEOREM 2.4.4. *For a given* $U \in \mathfrak{U}$ *these three conditions are equivalent:*

(2.4.41) U *has a representation* (2.3.1*) *with* $K(s, t)$ *measurable and* $K(s, .)$ *almost separably-valued in* $BV(T)$,

(2.4.42) U *has a representation* (2.3.1) *with* $K(s, .)$ *almost separably-valued in* $C^*(T)$,

(2.4.43) *there is a representation* $U(\phi) = \int_S \nu_s\phi(s)d\alpha$ *with* $\nu_\cdot \in \mathfrak{A}^\infty(S)[C^*(T)]$.

Each of these implies that

(2.4.44) $U'(\phi) = dU(\phi)/dt$ *can be written as* $U'(\phi) = \int_S \psi_s\phi(s)d\alpha$ *with* $\psi_\cdot \in \mathfrak{A}^\infty(S)[L(T)]$, *and*

(2.4.45) U' *has a representation* $U'(\phi) = \int_S K'(s, t)\phi(s)d\alpha$ *where* $K'(s, t)$ *is measurable,* $K'(s, .)$ *is defined to* $L(T)$, *and for almost every* s, $K'(s, t) = dK(s, t)/dt$ *a.e. in* T, $K(s, t)$ *being any kernel satisfying* (2.4.41) *or* (2.4.42).

If one of the conditions (2.4.41)–(2.4.43) *is satisfied, then* $|U'| = \text{ess. sup.}_s \int_T |K'(s, t)|\,d\beta$.

Perhaps the only remark needed here is this. Suppose (2.4.41) holds and let $K'(s, t)$ be the kernel given in Theorem 2.4.3. Then for almost every s we have $K'(s, t) = dK(s, t)/dt$ a.e. in T, where $K(s, t)$ is the kernel assumed to satisfy (2.4.41). But if $K_1(s, t)$ is any other kernel satisfying (2.4.41) or (2.4.42), then $K_1(s, t) = K(s, t)$ is an identity in t for almost all s, by Theorem 2.3.8. Thus $K'(s, t)$ fulfills every part of (2.4.45).

B. **Operations U' to $L(T)$ of the form $U'(\phi) = \int_S K'(s, t)\phi(s)d\alpha$.** Thanks to Theorem 2.4.4 several conditions can be given each of which is necessary and sufficient that an operator U' from $L(S)$ to $L(T)$ be representable in the form (2.4.02). In precise terms we have

THEOREM 2.4.5. *Let U' be an operation from $L(S)$ to $L(T)$. The following four conditions are equivalent*:

(2.4.51) *U' can be written as*

$$U'(\phi) = \int_S K'(s, t)\phi(s)d\alpha$$

where K' is measurable and ess. sup.$_s \int_T |K'(s, t)|\, d\beta < \infty$;

(2.4.52) *U' can be written as*

$$U'(\phi) = \int_S \psi_s \phi(s)d\alpha$$

where $\psi_\cdot \in \mathfrak{B}^\infty(S)[L(T)]$;

(2.4.53) *among the representations*

$$U'(\phi) = \frac{d}{dt}\int_S H(s, t)\phi(s)d\alpha$$

there is one with H measurable and $H(s, \cdot)$ essentially separably-valued in $BV(T)$,

(2.4.54) *among those $U \in \mathfrak{U}$ for which $U'(\phi) = dU(\phi)/dt$ there is one having representation $U(\phi) = \int_S \nu_s \phi(s)d\alpha$ where $\nu_\cdot \in \mathfrak{B}^\infty(S)[C^*(T)]$.*

Under condition (2.4.51) *the norm of U' is* $|U'| =$ ess. sup.$_s \int_T |K'(s, t)|\, d\beta$.

From Theorem 2.4.4 it is seen that (2.4.53) and (2.4.54) are equivalent and that either implies both (2.4.51) and (2.4.52). Conditions (2.4.51) and (2.4.52) are equivalent in view of Theorems 1.3.5 and 1.3.6.† To complete the proof it is sufficient to show that (2.4.52) implies (2.4.54). Let I be the operator from $L(T)$ to $AC(T)$ assigning to each ψ its indefinite integral. Setting

† Thus when T is abstract (2.4.51) and (2.4.52) are still equivalent, and the norm of U' is ess sup$_s \int_T |K'(s, t)|\, d\beta$ in case (2.4.51) is satisfied. The second statement was established by Vulich [42] for real intervals S and T.

$V(\phi)=I(U'(\phi))$ it is evident that $V \in \mathfrak{U}$ and that $U'(\phi)=dV(\phi)/dt$. Moreover $U(\phi)=I(\int_S \psi_s \phi(s)d\alpha)=\int_S \nu_s \phi(s)d\alpha$, where $\nu_s \equiv I(\psi_s)$ and $\nu_\cdot \in \mathfrak{A}^\infty(S)[AC(T)]$ since $\psi_\cdot \in \mathfrak{A}^\infty(S)[L(T)]$. Thus (2.4.54) is satisfied.

Two comments on condition (2.4.51) may be made. First, the requirement that ess. $\sup._s \int_T |K'(s,\ t)|d\beta$ be finite may be replaced by the equivalent one that $K'(s,\ .)$ be essentially defined to $L(T)$ and that ess. $\sup._s |\int_T K'(s,\ t)\psi'(t)d\beta|$ be finite for each $\psi'(.) \in L^\infty(T)$. Secondly, if S is Euclidean, K' is measurable, and $U'(\phi)=\int_S K'(s,\ t)\phi(s)d\alpha$ is an operation from $L(S)$ to $L(T)$, then necessarily ess. $\sup. \int_T |K'(s,\ t)|d\beta < \infty$, as Kantorovitch and Vulich have shown [24, p. 146].

C. **Operations to $AC(T)$ given by almost separably-valued kernels.** Among the operations to $C^*(T)$ that are representable by almost separably-valued kernels there is the proper subclass consisting of those having their range in $AC(T)$. This subclass is characterized in the following theorem.

THEOREM 2.4.6. *For a given $U \in \mathfrak{U}$ these five conditions are equivalent:*

(2.4.61) $U(\phi)=\int_S K(s,\ t)\phi(s)d\alpha$ *where K is measurable, $K(s,\ .)$ is almost separably-valued in $BV(T)$, and $\int_E K(s,\ t)d\alpha \in AC(T)$ for every $E \in \mathcal{E}_B$,*

(2.4.62) $U(\phi)=\int_S K(s,\ t)\phi(s)d\alpha$ *where $K(s,\ .)$ is essentially defined to $AC(T)$,*

(2.4.63) $U(\phi)=\int_S \nu_s \phi(s)d\alpha$ *with $\nu_\cdot \in \mathfrak{A}^\infty(S)[AC(T)]$,*

(2.4.64) $U(L(S)) \subset AC(T)$ *and $U'(\phi)=\int_S \psi_s \phi(s)d\alpha$ with $\psi_\cdot \in \mathfrak{A}^\infty(S)[L(T)]$.*

(2.4.65) $U(L(S)) \subset AC(T)$ *and $U'(\phi)=\int_S K'(s,\ t)\phi(s)d\alpha$ where K' is measurable and* ess. $\sup._s \int_T |K'(s,\ t)|d\beta < \infty$.

A kernel K satisfies (2.4.61) *if and only if it satisfies* (2.4.62); *if it satisfies either and if K' fulfills* (2.4.65), *then*

$$|U| = |U'| = \operatorname*{ess.\,sup.}_s \int_T |K'(s,\ t)|\,d\beta = \operatorname*{ess.\,sup.}_s [\operatorname{var}_t K(s,\ t)]$$

$$= \operatorname*{ess.\,sup.}_s \int_T \left|\frac{d}{dt}K(s,\ t)\right| d\beta$$

and, for almost every s, $K'(s,\ t)=dK(s,\ t)/dt$ a.e. in T.

Suppose (2.4.61) holds for a given kernel K. Theorem 2.4.2 then asserts that $U(\phi)=\int_S \nu_s \phi(s)d\alpha$ where $\nu_\cdot \in \mathfrak{A}^\infty(S)[C^*(T)]$ and $\nu_s = K(s,\ .)$ a.e. in S. From Theorem 2.4.1 and the assumption that $U(L(S)) \subset AC(T)$ it follows that $\nu_\cdot \in \mathfrak{A}^\infty(S)[AC(T)]$ and hence $K(s,\ .)$ is essentially defined to $AC(T)$. Thus if K fulfills (2.4.61) it also fulfills (2.4.62). When K satisfies (2.4.62) it follows from Theorems 2.4.2. and 2.3.8 that (2.4.63) holds and that K satisfies (2.4.61). Condition (2.4.64) follows easily from (2.3.63); it is sufficient to

recall Theorem 1.2.9 and the proof of Theorem 2.4.3. The equivalence of (2.4.64) and (2.4.65) is included in Theorem 2.4.5, and it is obvious that (2.4.65) implies (2.4.63). Finally, (2.4.61) results from (2.4.63) by Theorem 2.4.4. This completes the proof of the equivalence of the conditions (2.4.61)–(2.4.65) and the fact that K fulfills (2.4.61) if and only if it fulfills (2.4.62). The rest of the theorem can be inferred from Theorems 2.4.4 and 2.4.5.

THEOREM 2.4.7. *Let U be in $\mathfrak{U}$ and $U(\phi)=\int_S H(s, t)\phi(s)d\alpha$ where $H(s, .)$ is essentially defined to $C^*(T)$. Suppose any one of the conditions* (2.4.61)–(2.4.65) *is satisfied. If $K(s, t)$ is any kernel fulfilling* (2.4.61) *or* (2.4.62), *then $H(s, .)=K(s, .)$ in $C^*(T)$ for almost all s. Hence $H(s, .)$ is essentially defined to $AC(T)$.*

This comes immediately from Theorems 2.3.8, 2.4.2, and 2.4.6.

By considering two simple examples a few comments can be made on the preceding theorems. Set $S=T=[0, 1]$ and let U assign to each $\phi \in L(S)$ its indefinite integral. It is easily verified that $U(\phi)=\int_S H(s, t)\phi(s)d\alpha$ where $H(s, t)=0$ if $s>t$ and $H(s, t)=1$ if $s\leqq t$. From this and Theorem 2.4.7 there follows the well known fact that the identity operation I from $L(S)$ to $L(S)$ cannot be represented as $I(\phi)=\phi=\int_S K'(s, t)\phi(s)d\alpha$ where $K'(s, t)$ is measurable. For here U' coincides with I, and if I were so representable (i) and (ii) of (2.4.65) would be fulfilled; it would then result that $H(s, .)$ is essentially defined to $AC(T)$, which is clearly false. This example also shows that those operators from $L(S)$ to $AC(T)$ defined by almost separably-valued kernels form a proper subclass of all the operators on $L(S)$ to $AC(T)$. Consequently the set of all operations from $L(S)$ to $C^*(T)$, $S=T=[0, 1]$, is larger than the subclass consisting of those having representations by means of almost separably-valued kernels.

The operators characterized in Theorem 2.4.6 form an even smaller class. For an example it is enough to consider $U(\phi)=\int_S \eta(t)\xi(s)\phi(s)d\alpha$ where $S=T=[0, 1]$, $\eta(.) \in C^*(T)-AC(T)$, and $\xi(.)$ is bounded, measurable, and has $\infty >$ ess. sup. $|\xi(s)|>0$. Here $AC(T) \supset U(L(S))$, yet $H(s, .)\equiv\xi(s)\eta(.)$ is separably-valued and is in $\mathfrak{A}^\infty(S)[C^*(T)]$.

These two examples also prove that no one of the conditions (2.4.61)–(2.4.65) remains equivalent to the others if in that particular condition either $AC(T)$ is replaced by $C^*(T)$ or the assumption that $H(s, .)$ is almost separably-valued is dropped.

PART 5. THE ITERATIVE INTEGRALS ASSOCIATED WITH CERTAIN KERNELS

Before ending Chapter II we should like to make a few remarks concerning the existence and equality of the iterative integrals associated with some

of the kernels discussed in Parts 3 and 4. The following result, already established in Theorem 2.3.7, is an illustration.

THEOREM 2.5.1. *Suppose U is in* $\mathfrak{U}$ *and* $U(\phi)=\int_S H(s,t)\phi(s)d\alpha$ *where* (i) $H(s,t)$ *is measurable and* (ii) $H(s,t)$ *is everywhere continuous on the right in* t. *Then the two mixed integrals*

$$\mathrm{S}\int_T \gamma(t)d\left\{\int_S H(s,t)\phi(s)d\alpha\right\}, \qquad \int_S\left\{\mathrm{S}\int_T \gamma(t)dH(s,t)\right\}\phi(s)d\alpha$$

exist and are equal when $\gamma(.)\,\varepsilon\, C(T)$ *and* $\phi(.)\,\varepsilon\, L(S)$.

Another result of the same sort is

THEOREM 2.5.2. *Theorem* 2.5.1 *remains true if in place of* (ii) *it is assumed that* $\mathrm{var}_t H(s,t)<\infty$ *for almost every* s.

Let $K(s,t)$ be the kernel given in the conclusions of Theorem 2.3.6; the above two integrals then necessarily exist and coincide when H is replaced by K. In addition, $\int_S H(s,t)\phi(s)d\alpha=\int_S K(s,t)\phi(s)d\alpha$, so that

$$\mathrm{S}\int_T \gamma(t)d\left\{\int_S H(s,t)\phi(s)d\alpha\right\}=\int_S\left\{\mathrm{S}\int_T \gamma(t)dK(s,t)\right\}\phi(s)d\alpha.$$

On the other hand, for almost every s the two functions $H(s,t)$ and $K(s,t)$ are of bounded variation with respect to t and coincide a.e. in T. This implies that $\mathrm{S}\int_T\gamma(t)dH(s,t)=\mathrm{S}\int_T\gamma(t)dK(s,t)$ for every $\gamma(.)\,\varepsilon\, C(T)$ and almost every s, and hence

$$\int_S\left\{\mathrm{S}\int_T \gamma(t)dH(s,t)\right\}\phi(s)d\alpha=\int_S\left\{\mathrm{S}\int_T \gamma(t)dK(s,t)\right\}\phi(s)d\alpha.$$

For almost separably-valued kernels a stronger conclusion can be made.

THEOREM 2.5.3. *If, in Theorem* 2.5.1, (ii) *is replaced by the assumption that* $H(s,.)$ *is almost separably-valued in* $BV(T)$, *then the two mixed integrals* (*Lebesgue and Radon-Stieltjes*)

$$\int_S \phi(s)\left\{\int_Q H(s,t)dF(P)\right\}d\alpha, \qquad \int_Q\left\{\int_S H(s,t)\phi(s)d\alpha\right\}dF(P)$$

exist and are equal for each $\phi\,\varepsilon\, L(S)$ *and each* $F(P)$ *which is bounded and additive over all subsets* P *of the space* Q *consisting of all partitions of* T.

From Theorem 2.4.2 the function $\nu_s\equiv H(s,.)$ is in $\mathfrak{L}^{\infty}(S)[C^*(T)]$ and $U(\phi)=\int_S\nu_s\phi(s)d\alpha$. For each linear functional $f(\nu)$ over $C^*(T)$ we then have $f(U(\phi))=\int_S f(\nu_s)\phi(s)d\alpha$. Since each $F(P)$ defines such a functional by means of the Radon-Stieltjes integral [20] and since $\nu_s=H(s,.)$ in $C^*(T)$ for

almost every s, it follows that $\int_Q \{\int_S H(s, t)\phi(s)d\alpha\} dF(P) = f(U(\phi))$ $= \int_S \{\int_Q H(s, t)dF(P)\}\phi(s)d\alpha$.

THEOREM 2.5.4. *From the assumptions of Theorem* 2.5.3 *it also follows that*

$$\int_T \psi(t)\left[\frac{d}{dt}\int_S H(s, t)\phi(s)d\alpha\right]dt = \int_S \phi(s)\left[\int_T \psi(t)\left(\frac{d}{dt}H(s, t)\right)dt\right]d\alpha$$

whenever $\phi \in L(S)$ *and* $\psi \in L^\infty(T)$. *Hence*

$$\int_c^\tau \left[\frac{d}{dt}\int_E H(s, t)d\alpha\right]dt = \int_E \left[\int_c^\tau \left(\frac{d}{dt}H(s, t)\right)dt\right]d\alpha$$

for each $\tau \in T$ *and each* $E \in \mathcal{E}_B$.

Here the operation $U'(\phi) = (d/dt)\int_S H(s, t)\phi(s)d\alpha$ can be written as $U'(\phi) = \int_S \psi_s\phi(s)d\alpha$ where $\psi_\cdot \in \mathfrak{A}^\infty(S)[L(T)]$ and ψ_s is for almost every s the element $dH(s, t)/dt$ in $L(T)$. Thus $f(U'(\phi)) = \int_S f(\psi_s)\phi(s)d\alpha$ for every f in the adjoint of $L(T)$, so that the above two integrals exist and are equal for each $\phi \in L(S)$ and $\psi \in L^\infty(T)$.

CHAPTER III. RESTRICTED OPERATIONS TO ARBITRARY SPACES

PART 1. WEAKLY COMPLETELY CONTINUOUS AND COMPLETELY CONTINUOUS OPERATIONS

In this chapter the restrictions and emphasis are placed upon the operation rather than upon the range space. Only those operations U will be considered which map $L(S)$ into an arbitrary B-space X and fulfill one of the requirements (i) U is defined by an element of $\mathfrak{A}^\infty(S)[X]$, (ii) U is weakly c.c., or (iii) it is c.c. Our aim is to obtain properties and vector integral representations of the three classes of operators defined respectively by these conditions.

Part 1 contains the proof that every U satisfying (i) must map weakly compact sets into compact sets and also provides conditions necessary and sufficient that an operation which satisfies (i) also satisfy (ii) or (iii). When S is Euclidean we are able to be more precise; in Theorem 3.1.9 characteristic representations are given for the weakly c.c. and the c.c. operations from $L(S)$ to* X. In addition it is shown, without restricting S or T, that a mapping U is a c.c. operator from $L(S)$ to $L^q(T)$, $1 \leqq q \leqq \infty$, if and only if $U(\phi) = \int_S x(s)\phi(s)d\alpha$ where $x(.)$ is an essentially compact-valued element of $\mathfrak{A}^\infty(S)[L^q(T)]$.

* The representation to be given for c.c. operations was originally due to Gelfand [18], who established it for the case in which S is a bounded real interval.

A. Sufficient conditions for weak complete continuity and complete continuity. We begin by considering those operations generated by certain abstract integrals.

THEOREM 3.1.1. *Let $x(.)$ be in $\mathfrak{L}_0^\infty(S)[X, X^*]$. Then* (i) *if $x(.)$ is essentially weakly compact-valued, it is essentially bounded and the operation $U(\phi) = \int_S x(s)\phi(s)d\alpha$ is weakly c.c. from $L(S)$ to X; and* (ii) *if $x(.)$ is essentially compact-valued, it is in $\mathfrak{A}^\infty(S)[X]$ and U is c.c.*

Suppose $E_0 \in \mathcal{E}_0$ is such that $x(S-E_0)$ is weakly compact. Every weakly compact set being bounded, $x(s)$ obviously is essentially bounded. Moreover, $R(x(S-E_0))$ must be weakly compact, so that the point set sum $Y=x(S-E_0)+R(x(S-E_0))$ has the same property. From a theorem of Chmoulyan [39] it follows that the closed convex hull $C[Y]$ is necessarily weakly compact. Since $C[Y] \supset U(\phi)$ for every ϕ with $\|\phi\| \leqq 1$ by part (iii) of Theorem 1.2.5, clearly U is weakly c.c. If $x(S-E_0)$ satisfies the stronger condition of being compact, then $x(.)$ is essentially bounded and almost separably-valued, and is therefore in $\mathfrak{A}^\infty(S)[X]$ by Theorem 1.1.7 since $f(x(s))$ is in $L^\infty(S)$ for every $f \in X^*$. In addition the set Y is compact and so $C[Y]$ is compact, from a result of Mazur [28]. Again applying Theorem 1.2.5 it is seen that U is c.c.

The next theorem is one of the most important in the paper.

THEOREM 3.1.2. *For an otherwise arbitrary $x(s)$ suppose there exist an $E_0 \in \mathcal{E}_0$ and a c.l.m. $\Gamma \subset X^*$ such that $x(S-E_0)$ is separable and Γ is a determining manifold for $x(S-E_0)$. If $x(.) \in \mathfrak{L}^\infty(S)[X, \Gamma]$, then $x(.) \in \mathfrak{A}^\infty(S)[X]$. Moreover the following are true for $x(.)$ since they are true for any element of $\mathfrak{A}^\infty(S)[X]$:*

(I) *$U(\phi) = \int_S x(s)\phi(s)d\alpha$ defines a separable operation U from $L(S)$ to X that takes weakly compact sets into compact sets and has $|U| = \text{ess. sup. } \|x(s)\|$,*

(II) *U is weakly c.c. if and only if $x(.)$ is essentially weakly compact-valued,*

(III) *U is c.c. if and only if $x(.)$ is essentially compact-valued.*

The first statement is given by Theorem 1.1.7. Now suppose $x(.)$ is any element of $\mathfrak{A}^\infty(S)[X]$. Since conclusions (II) and (III) follow immediately from Theorem 3.1.1 and (IV) of Theorem 1.2.10, only (I) remains to be proved. We thus wish to show that $U(\Phi)$ is compact when Φ is weakly compact and $x(.)$ is an arbitrary element in the class $\mathfrak{A}^\infty(S)[X]$.

Suppose $x(.)$ is almost countably-valued, that is,

(1) in $\mathcal{E}$ there are disjoint elements E_i, $i=0, 1, 2, \cdots$, such that $\sum_0^\infty E_i = S$, $E_0 \in \mathcal{E}_0$, and, for $i \geqq 1$, $x(s)$ has a constant value x_i on E_i.

For each $\phi \in L(S)$ we have $\sum_1^\infty |\int_{E_i}\phi(s)d\alpha| \leqq \|\phi\|$ and hence $P(\phi) = \{\int_{E_i}\phi(s)d\alpha\}$, $(i=1, 2, \cdots)$, is an operation from $L(S)$ to l. The image $P(\Phi)$ of the weakly compact set Φ must then be weakly compact. Since in l compactness and weak compactness are equivalent properties, the set $P(\Phi)$ is actually compact. Now consider the bounded sequence $\{x_i\}$. The mapping $Q(\psi) = \sum_1^\infty x_i\psi_i$ is defined from l to X and is an operation, so that $Q(P(\Phi))$ is compact in X. But for each $\phi \in L(S)$ the equality $U(\phi) = \sum_1^\infty \int_{E_i}x(s)\phi(s)d\alpha = \sum_1^\infty x_i\int_{E_i}\phi(s)d\alpha$ holds, and hence $U(\phi) = Q(P(\phi))$. Thus U takes weakly compact sets into compact sets when $x(.)$ satisfies (1).

Now let $x(.)$ in $\mathfrak{A}^\infty(S)[X]$ be arbitrary and let K be a bound for the weakly compact set Φ. From the measurability of $x(.)$ it follows [31, Corollary 1.12] that for each $\epsilon>0$ a measurable function $x'(s)$ exists satisfying both (1) and

(2) ess. sup. $\|x(s)-x'(s)\| < \epsilon/2K$.

From this fact we can conclude that $U(\Phi)$ is totally bounded and hence is compact. For $x'(.) \in \mathfrak{A}^\infty(S)[X]$ as a consequence of (2), so that by the preceding case (1) implies that $U'(\phi) = \int_S x'(s)\phi(s)d\alpha$ is an operation mapping Φ into a compact set. Thus $U'(\Phi)$ is totally bounded and can be covered by a finite number of spheres having centers x_i', $i=1, 2, \cdots, m$, and radii each less than $\epsilon/2$. Since

$$\|U(\phi) - U'(\phi)\| \leqq \text{ess. sup.} \|x(s) - x'(s)\| \cdot \|\phi\|, \qquad \phi \in L(S),$$

we have $\|U(\phi)-U'(\phi)\| < \epsilon/2$ for $\phi \in \Phi$. This implies that for each $\phi \in \Phi$ the point $U(\phi)$ is within ϵ of one of the centers x_i'. Thus $U(\Phi)$ is totally bounded.

The next four theorems are immediate corollaries to Theorem 3.1.2.

THEOREM 3.1.3. *A given measurable function $x(s)$ is essentially (weakly) compact-valued if and only if $U(\phi) = \int_S x(s)\phi(s)d\alpha$ is defined and (weakly) c.c. from $L(S)$ to X.*

THEOREM 3.1.4. *A necessary and sufficient condition that $\{x_n\} \subset X$ be a (weakly) compact sequence is that $\{x_n\}$ is bounded and the operation $U(\phi) = \sum_n \phi_n x_n$, $\phi = \{\phi_n\}$, from l to X is (weakly) c.c.**

THEOREM 3.1.5. *U is defined and (weakly) c.c. from l to X if and only if if it can be written as $U(\phi) = \sum_n \phi_n x_n$ where $\{x_n\}$ is (weakly) compact in X. The norm of U is† $\sup \|x_n\|$.*

THEOREM 3.1.6. *If $x(.) \in \mathfrak{A}^\infty(S)[X]$ is essentially weakly compact-valued, then $U(\phi) = \int_S x(s)\phi(s)d\alpha$ is weakly c.c. from $L(S)$ to X and takes weakly com-*

* The two propositions given in Theorem 3.1.4 are equivalent respectively to those of Chmoulyan and Mazur used in proving Theorem 3.1.1.

† The statement concerning c.c. operations that is included in this theorem is due to Gelfand [18].

pact sets into compact sets. Hence, if $X=L(S)$, the iterated operation $U^n(\phi)$ is defined and c.c. from $L(S)$ to $L(S)$ for $n\geqq 2$, which implies that the fixed points of U form a finite-dimensional c.l.m. in $L(S)$.

From (I) of Theorem 3.1.2 and Theorems 2.1.8 and 2.1.9 we also have

THEOREM 3.1.7. *If X is either a separable adjoint space or has a weakly compact unit sphere, then*

(i) *U is an operation defined and c.c. from $L(S)$ to X if and only if $U(\phi)=\int_S x(s)\phi(s)d\alpha$ where $x(.)$ is an essentially compact-valued element of $\mathfrak{N}^\infty(S)[X]$.*

If X has a weakly compact unit sphere, then

(ii) *U is an operation defined, separable, and weakly c.c. from $L(S)$ to X if and only if $U(\phi)=\int_S x(s)\phi(s)d\alpha$ where $x(.)$ is an essentially weakly compact-valued element of $\mathfrak{N}^\infty(S)[X]$, and*

(iii) *every separable operation from $L(S)$ to X takes weakly compact sets into compact sets.*

When X is a separable adjoint space, statements (ii) *and* (iii) *are true with the word "separable" deleted.*

The space $L^q(T)$ having a weakly compact unit sphere in case $1<q<\infty$, Theorem 3.1.7 yields

THEOREM 3.1.8. *When $X=L^q(T)$, $1<q<\infty$, conclusions* (i)–(iii) *of Theorem 3.1.7 hold.*

B. **Criteria for weak complete continuity and complete continuity when S is Euclidean.** The theorem below solves for Euclidean S the problems of characterizing the weakly c.c. and the c.c. operations from $L(S)$ to an arbitrary X.

THEOREM 3.1.9. *Let S be a Euclidean interval, finite or infinite, and let α be Lebesgue measure. An operation U is defined and (weakly) c.c. from $L(S)$ to X if and only if there exists in $\mathfrak{N}^\infty(S)[X]$ an essentially (weakly) compact-valued $x(.)$ such that $U(\phi)=\int_S x(s)\phi(s)d\alpha$. The norm of U is* ess. sup. $\|x(s)\|$.

By reason of Theorem 3.1.2 and the fact that complete continuity implies weak complete continuity, it suffices to prove that every weakly c.c. operation U from $L(S)$ to X has a representation $U(\phi)=\int_S x(s)\phi(s)d\alpha$ where $x(.)\in\mathfrak{N}^\infty(S)[X]$. Suppose U to be such an operation and S to be bounded. Consider a fixed point s in the interior S^0 of S and let $\{I_n'\}$ be some sequence of nondegenerate cubes such that $S^0\supset I_n'\ni s$ and $\lim_n\alpha(I_n')=0$. If $x_n'=U(\phi_n')$ where $\phi_n'=\phi_{I_n'}/\alpha(I_n')$ then since $\|\phi_n'\|\leqq 1$ and U is weakly c.c. it follows that $\{x_n'\}$ contains a subsequence $\{x_n\}$ converging weakly to an element $x(s)$ of X. From this convergence property it is evident for any $f\in X^*$ that

$f(x(s))=\lim_n f(x_n)=\lim_n f(X_{I_n})/\alpha(I_n)$ where X_R is the image under U of the characteristic function of the figure R. Since $\{I_n\}$ closes down on s, this implies that if $f(X_R)$ is differentiable at s its derivative there has the value $f(x(s))$. Since $\alpha(S-S^0)=0$, the function $x(s)$ is thus defined a.e. in S and has the property that for each $f \in X^*$ the numerical additive Lipschitzean function $f(X_R)$ has its derivative coinciding with $f(x(s))$ wherever this derivative exists. At this point we make the observation that since $L(S)$ is separable and U is continuous, there is no loss of generality in assuming X to be separable. We now have the following situation: (i) X_R is an additive Lipschitzean function defined to X from the figures lying in the bounded Euclidean interval S, (ii) $x(s)$ is a point function defined a.e. in S and having its values in the separable space X, and (iii) for each $f \in X^*$ the numerical function $f(X_R)$ is differentiable a.e. in S to the value $f(x(s))$. From Theorem 2.7 of [32] it follows immediately that $x(.)$ is measurable and integrable and that X_R is differentiable a.e. in S to the value $x(s)$. Since X_R is additive and Lipschitzean, we see that $X_R=\int_R x(s)d\alpha$ and $x(.) \in \mathfrak{A}^\infty(S)[X]$. Thus $U(\phi)=\int_S x(s)\phi(s)d\alpha$ where $x(.) \in \mathfrak{A}^\infty(S)[X]$.

In case S is unbounded let $\{S_i\}$ be a decomposition of S into bounded intervals. By the preceding case there is for each i an $x_i(s)$ defined a.e. in S_i and such that $U(\phi)=\int_{S_i} x_i(s)\phi(s)d\alpha$ for each $\phi(.) \in L(S)$ that vanishes a.e. in $S-S_i$; moveover, $x_i(s)$ is measurable in S and $|U|=\text{ess. sup. } \|x_i(s)\|$. Let $x(s)=x_i(s)$ when $s \in S_i$ and $x_i(s)$ is defined at s. Clearly $x(.) \in \mathfrak{A}^\infty(S)[X]$, and since any simple function $\phi(s)$ can be written as $\phi(s)=\sum_1^n \phi_i(s)$, where each $\phi_i(s)$ vanishes outside of S_i, we also have $U(\phi)=\sum_1^n U(\phi_i)=\sum_1^n \int_{S_i} x_i(s)\phi_i(s)d\alpha$ $=\sum_1^n \int_S x(s)\phi_i(s)d\alpha=\int_S x(s)\phi(s)d\alpha$ whenever $\phi(s)$ is simple. This of course implies that $U(\phi)=\int_S x(s)\phi(s)d\alpha$ for every $\phi \in L(S)$.

COROLLARY. *When S is Euclidean and α is Lebesgue measure, any weakly c.c. operation U from $L(S)$ to X has the property of taking weakly compact sets into compact sets. Hence if $X=L(S)$ the nth iterate U^n is c.c. for $n \geqq 2$ and thus U has for its fixed points a finite-dimensional c.l.m.*

C. **Abstract representations of completely continuous operations to $L^q(T)$, $1 \leqq q \leqq \infty$.** We turn now to the proof that the characterization of c.c. operators given in the last theorem is also true for unrestricted S provided that X is an $L^q(T)$ space.

THEOREM 3.1.10. *An operation U is defined and c.c. from $L(S)$ to $L^q(T)$, $1 \leqq q \leqq \infty$, if and only if $U(\phi)=\int_S x(s)\phi(s)d\alpha$ where $x(.)$ is an essentially compact-valued element of $\mathfrak{A}^\infty(S)[L^q(T)]$. The norm of U is $|U|=$ ess. sup. $\|x(s)\|$.*

The case in which $1<q<\infty$ has been covered in Theorem 3.1.8. In virtue

of Theorem 3.1.2 all that is required then is the proof that each c.c. operation from $L(S)$ to $L^q(T)$, $q=1$ or ∞, can be represented by some element of $\mathfrak{M}^\infty(S)[L^q(T)]$. The case $q=1$ will be considered first.

Let π stand for a finite number of disjoint measurable sets Δ in T such that $\infty > \beta(\Delta) > 0$, $\Delta \in \pi$. If $\pi=(\Delta)$ and $\pi'=(\Delta')$ are two such partitionings, define $\pi\pi'$ to be the partitioning consisting of those $\Delta\Delta'$ with $\beta(\Delta\Delta')>0$. We shall use the relation $\pi \subset \pi'$ to mean that for every $\Delta \in \pi$ there is a $\Delta' \in \pi'$ with $\beta(\Delta-\Delta')=0$. It is clear then that $\pi\pi' \subset \pi$. If x and $x(\pi)$ are points in a B-space, the statement

$$\lim_{\pi} x(\pi) = x \tag{1}$$

means that for every $\epsilon>0$ there is a π_ϵ such that $\|x(\pi)-x\|<\epsilon$ if $\pi \subset \pi_\epsilon$. If (1) holds, there is a sequence π_n with $x(\pi_n)\to x$. Similarly we say that $x(\pi)$ is convergent if

$$\|x(\pi) - x(\pi')\| < \epsilon \qquad \text{for } \pi, \pi' \subset \pi_\epsilon.$$

If (1) holds, $x(\pi)$ is convergent and if $x(\pi)$ is convergent there is a unique x for which (1) is true.

Let Y_π be the c.l.m. in $L(T)$ consisting of those elements which are constant on each of the sets $\Delta \in \pi$. Thus

$$Y_{\pi'} \subset Y_\pi \qquad \text{if } \pi \subset \pi'. \tag{2}$$

Let ψ_Δ be the characteristic function of Δ and U_π the projection of $L(T)$ into Y_π which is defined by

$$U_\pi(\psi) = \sum_{\Delta \in \pi} \psi_\Delta \frac{\int_\Delta \psi(t)d\beta}{\beta(\Delta)}. \tag{3}$$

With $U_\pi(\psi)$ thus defined we have

$$|U_\pi| = 1, \qquad U_\pi(\psi) = \psi \qquad \text{for } \psi \in Y_\pi. \tag{4}$$

From (2) and (4) we see that for a simple function ψ', say $\psi' \in Y_{\pi'}$, we have $U_\pi(\psi')=\psi'$ for $\pi \subset \pi'$ and thus for simple functions

$$\lim_{\pi} U_\pi(\psi) = \psi. \tag{5}$$

Since the simple functions are dense in $L(T)$ and $|U_\pi|=1$, it follows that (5) holds for every $\psi \in L(T)$. Moreover (5) holds uniformly over any compact set. To see this we take points ψ_i, $i=1, 2, \cdots, n_\epsilon$, in a compact set Ψ such that every point in Ψ is within ϵ distance of some ψ_i. There is then a π_ϵ such that

$$\|U_\pi(\psi_i) - \psi_i\| < \epsilon, \qquad i = 1, 2, \cdots, n_\epsilon, \ \pi \subset \pi_\epsilon,$$

and thus

$$\|U_\pi(\psi) - \psi\| < 3\epsilon \qquad \text{for } \psi \in \Psi, \ \pi \subset \pi_\epsilon.$$

Hence if U is c.c. from $L(S)$ into $L(T)$ we have

$$\lim_\pi \ | U_\pi U - U | = 0.$$

There is then a sequence π_n such that

$$\lim_n \ | U_{\pi_n} U - U | = 0, \qquad \lim_{m,n} \ | U_{\pi_m} U - U_{\pi_n} U | = 0. \tag{6}$$

For a c.c. operation $\psi = U(\phi)$ on $L(S)$ to $L(T)$ we have (see (3))

$$U_\pi(U(\phi)) = \sum_{\Delta \in \pi} f_\Delta(\phi)\psi_\Delta,$$

where $f_\Delta(\phi) = \int_\Delta \psi(t)d\beta/\beta(\Delta)$ is a linear functional on $L(S)$. Thus there are elements $\mu_\Delta(.) \in L^\infty(S)$ such that

$$U_\pi(U(\phi)) = \sum_{\Delta \in \pi} \psi_\Delta \int_S \mu_\Delta(s)\phi(s)d\alpha = \int_S x_\pi(s)\phi(s)d\alpha \tag{7}$$

where $x_\pi(.)$ is the abstract function $x_\pi(s) = \sum_{\Delta \in \pi} \psi_\Delta \mu_\Delta(s)$. Writing $x_n(s)$ for $x_{\pi_n}(s)$, it is clear that $x_n(.) \in \mathfrak{A}^\infty(S)[L(T)]$. From (6), (7), and Theorem 1.1.7

$$\lim_{m,n} \text{ess. sup.} \|x_m(s) - x_n(s)\| = 0,$$

and hence $x(s) = \lim_m x_m(s)$ is defined a.e. and $x(.) \in \mathfrak{A}^\infty(S)[L(T)]$, with

$$\lim_m \text{ess. sup.} \|x_m(s) - x(s)\| = 0. \tag{8}$$

Thus $U'(\phi) = \int_S x(s)\phi(s)d\alpha$ is an operation from $L(S)$ to $L(T)$ and (8) shows that $\lim_m | U_{\pi_m} U - U' | = 0$. Combining this with (6) we have $U = U'$ and hence $U(\phi) = \int_S x(s)\phi(s)d\alpha$ where $x(.) \in \mathfrak{A}^\infty(S)[L(T)]$.

The proof for $q = \infty$ is somewhat similar. Let U be c.c. from $L(S)$ to $L^\infty(T)$ and suppose to begin with that $\beta(T) < \infty$. The space $L^\infty(T)$ can then be considered as a linear subset Y of $L(T)$, and the "identity" transformation $I(\mu) = \mu$ is a continuous 1-1 operation taking $L^\infty(T)$ into Y. Since convergence in the mean implies convergence in measure, the operation I is a homeomorphism on each closed compact set in $L^\infty(T)$. Set $V(\phi) = I(U(\phi))$ and let Φ be the unit sphere in $L(S)$. The closure M of $U(\Phi)$ is compact and hence $I(M)$ is closed and compact since I is necessarily a homeomorphism on M. In addition the compactness of $U(\Phi)$ implies that $I(M)$ is the closure of

$V(\Phi)$. The operation V to $L(T)$ clearly being c.c., we have by the preceding case that $V(\phi)=\int_S y(s)\phi(s)d\alpha$ where $y(.)\in\mathfrak{A}^\infty(S)[L(T)]$. From Theorem 1.2.10 it follows that $y(.)$ is essentially defined to the closure of $V(\Phi)$, that is, to $I(M)$. But M is closed and compact so that $I(M)$ maps into the closed compact set M under the inverse of I. Hence $x(s)=I^{-1}(y(s))$ is essentially compact-valued. It is therefore essentially bounded and almost separably-valued. To show that $x(.)\in\mathfrak{A}^\infty(S)[L^\infty(T)]$ it is enough to prove that $g_\psi(x(s))$ is measurable for each linear functional g_ψ over $L^\infty(T)$ defined by an element ψ of $L(T)$. Let $F\in\mathcal{F}_B$ be arbitrary. If g_F is the functional defined by the characteristic function of F, we have $g_F(x(s))=\int_F K(s,t)d\beta$ where $K(s,.)=x(s)$ in $L^\infty(T)$ for each s. But then $K(s,.)=y(s)$ in $L(T)$ for each s, and therefore $g_F(x(s))=f_F(y(s))$ where f_F is the linear functional over $L(T)$ defined by the characteristic function of F. Since the vector function $y(.)$ is measurable, we conclude that $g_F(x(s))$ is measurable for each $F\in\mathcal{F}_B$. The extension to g_ψ where ψ is a simple function is obvious, and, since simple functions are dense in $L(T)$, $g_\psi(x(s))$ must be measurable for every $\psi\in L(T)$. Thus $x(.)\in\mathfrak{A}^\infty(S)[L(T)]$. Let $U'(\phi)=\int_S x(s)\phi(s)d\alpha$. Then $g_F(U'(\phi))=\int_S g_F(x(s))\phi(s)d\alpha$ for each $F\in\mathcal{F}_B$. We also have $f_F(V(\phi))=\int_S f_F(y(s))\phi(s)d\alpha$. Since $f_F(y(s))=g_F(x(s))$ for every s and F, clearly $g_F(U'(\phi))=f_F(V(\phi))$. But obviously $f_F(V(\phi))=g_F(U(\phi))$, I being the "identity." Hence $g_F(U'(\phi))=g_F(U(\phi))$ for every $\phi\in L(S)$ and $F\in\mathcal{F}_B$. Since the g_F's are a total set of functionals over $L^\infty(T)$, the last equality yields $U'(\phi)=U(\phi)$, that is, $U(\phi)=\int_S x(s)\phi(s)d\alpha$ where $x(.)\in\mathfrak{A}^\infty(S)[L^\infty(T)]$. This concludes the case $\beta(T)<\infty$.

In the general case let $\{T_i\}$ be a decomposition of T and let X_n be the c.l.m. in $L^\infty(T)$ composed of all the elements which vanish a.e. in $T-\sum_1^n T_i$. If Π_n assigns to each μ in $L^\infty(T)$ the element of X_n which coincides with $\mu(t)$ a.e. in $\sum_1^n T_i$, clearly Π_n is an operation in $L^\infty(T)$ with $|\Pi_n|=1$. Suppose U is c.c. from $L(S)$ to $L^\infty(T)$. Set $V_n(\phi)=\Pi_n(U(\phi))$. Then V_n is c.c. to X_n, and since X_n is equivalent to $L^\infty(\sum_1^n T_i)$ where $\beta(\sum_1^n T_i)<\infty$ it follows from the preceding case that $V_n(\phi)=\int_S x_n(s)\phi(s)d\alpha$ where $x_n(.)\in\mathfrak{A}^\infty(S)[X_n]$. According to Theorem 1.2.10 $x_n(.)$ is essentially defined to the closure of $V_n(\Phi)$. Since $V_n(\Phi)=\Pi_n(U(\Phi))$ where $U(\Phi)$ is compact, we can infer that

$$x_n(.)\text{ is essentially defined to }\Pi_n(M),\qquad n=1,2,\cdots, \tag{9}$$

where M is the bounded compact closure of $U(\Phi)$. Going back to the definition of Π_j it is clear that for any $\mu\in L^\infty(T)$

$$\Pi_h\Pi_i(\mu)=\Pi_j(\mu)\qquad\text{for } h,\ i\geqq j. \tag{10}$$

Thus when $j\leqq h, i$ we have $V_j(\phi)=\Pi_h\Pi_i(U(\phi))=\Pi_h(V_i(\phi))=\int_S\Pi_h(x_i(s))\phi(s)d\alpha$ where $\Pi_h(x_i(.))\in\mathfrak{A}^\infty(S)[L^\infty(T)]$. But V_j is also defined by the element $x_j(.)$

of $\mathfrak{A}^\infty(S)[L^\infty(T)]$. Hence

$$x_j(s) = \Pi_h(x_i(s)) \text{ a.e. in } S, \qquad \text{if } j \leqq h, i. \tag{11}$$

From (9) and (11) we can now draw this conclusion: there is a null set E_0 such that if $s \in S - E_0$ and h, i, j, n are any integers then

$$x_n(s) \in \Pi_n(M), \tag{12}$$

$$x_j(s) = \Pi_h(x_i(s)), \qquad \text{if } j \leqq h, i. \tag{13}$$

From (12) we also have

$$\|x_n(s)\| \leqq |U|, \qquad s \in S - E_0. \tag{14}$$

Fix s in $S - E_0$ and let $x_n = x_n(s)$. Due to (12) there is for each n an element μ_n in the closed compact set M such that $x_n = \Pi_n(\mu_n)$. If $i \geqq j$ we then have

$$\Pi_j(\mu_i) = \Pi_i\Pi_j(\mu_i) = \Pi_j\Pi_i(\mu_i) = \Pi_j(x_i)$$

by (10), so that

$$\Pi_j(\mu_i) = x_j \qquad \text{for } i \geqq j, \tag{15}$$

from (13). Now let x_0 be that measurable function of t which coincides over each T_n with x_n. In view of (14), x_0 is in $L^\infty(T)$, and from (13) and the definition of x_0 it is easily verified that

$$\Pi_j(x_0) = x_j, \qquad j = 1, 2, \cdots. \tag{16}$$

In the compact sequence $\{\mu_n\}$ choose a subsequence $\{\mu_{n_i}\}$ converging to $\mu_0 \in M$. Fixing j we have $\Pi_j(\mu_0) = \lim_i \Pi_j(\mu_{n_i})$ or, from (15), $\Pi_j(\mu_0) = \lim_i x_j = x_j$. Combining this with (16) yields

$$\Pi_j(x_0) = \Pi_j(\mu_0), \qquad j = 1, 2, \cdots,$$

which implies that $x_0 = \mu_0$. Thus $x_0 \in M$.

For each $s \in S - E_0$ define $x_0(s)$ to be this point x_0 and for $s \in E_0$ let $x_0(s)$ vanish. Since $x_0(.)$ is essentially defined to the compact set M, it is essentially bounded and almost separably-valued. To prove that $x_0(.) \in \mathfrak{A}^\infty(S)[L^\infty(T)]$ it is therefore sufficient to obtain measurability for $f_\psi(x_0(s))$ whenever f_ψ is a linear functional over $L^\infty(T)$ defined by an element ψ of $L(T)$. For a functional of this sort it is obvious that

$$f_\psi(\mu) = \lim_n \int_{\Sigma_1^n T_i} \psi(t)\mu(t)d\beta = \lim_n f_\psi(\Pi_n(\mu)) \tag{17}$$

for every $\mu \in L^\infty(T)$. This together with (16) yields the equality $f_\psi(x_0(s)) = \lim_n f_\psi(\Pi_n(x_0(s))) = \lim_n f_\psi(x_n(s))$ for almost all s. Since each $x_n(.)$ is measurable, $f_\psi(x_0(s))$ must therefore be measurable. Thus $x_0(.) \in \mathfrak{A}^\infty(S)[L^\infty(T)]$.

The final assertion is that $U(\phi)=\int_S x_0(s)\phi(s)d\alpha$. For (17) leads to the equality

$$f_\psi(U(\phi)) = \lim_n f_\psi(V_n(\phi)) = \lim_n \int_S f_\psi(x_n(s))\phi(s)d\alpha \tag{18}$$

for $\phi \in L(S)$ and $\psi \in L(T)$, and it was just seen that $f_\psi(x_0(s))=\lim_n f_\psi(x_n(s))$ a.e. in S for each ψ. Hence

$$f_\psi(U(\phi)) = \int_S f_\psi(x_0(s))\phi(s)d\alpha, \qquad \phi \in L(S), \psi \in L(T),$$

or $f_\psi(U(\phi))=f_\psi(\int_S x_0(s)\phi(s)d\alpha)$ since $x_0(.) \in \mathfrak{A}^\infty(S)[L^\infty(T)]$. The functionals f_ψ forming a total set over $L^\infty(T)$, we arrive at the conclusion that $U(\phi)=\int_S x_0(s)\phi(s)d\alpha$.

D. **Approximation theorems.** The next theorem provides an approximation criterion for weak complete continuity.

THEOREM 3.1.11. *An operation U is defined, separable, and weakly c.c. from $L(S)$ to X if there exists a sequence $x_n(s)$ such that*

(i) $x_n(.) \in \mathfrak{A}^\infty(S)[X]$, $n=1, 2, \cdots$,

(ii) $\lim_{m,n}$ ess. $\sup_s \|x_m(s)-x_n(s)\|=0$,

(iii) *each $x_n(.)$ is essentially weakly compact-valued,*

(iv) *U coincides with the operation $V(\phi)=\int_S x(s)\phi(s)d\alpha$ defined by the element $x(s)=\lim_n x_n(s)$ of $\mathfrak{A}^\infty(S)[X]$.*

If U is weakly c.c. from $L(S)$ to X, S is Euclidean, and α is Lebesgue measure, then elements $x(.)$, $x_n(.)$, $n=1, 2, \cdots$, exist in $\mathfrak{A}^\infty(S)[X]$ such that conditions (i)–(iv) *are satisfied. Each $x_n(.)$ may be chosen to have only a countable number of functional values.*

If a sequence $\{x_n(.)\}$ exists satisfying (i)–(iv) it is clear that $x(s)$ exists a.e. in S and $x(.)$ is in $\mathfrak{A}^\infty(S)[X]$, so that $V(\phi)=\int_S x(s)\phi(s)d\alpha$ is defined and linear from $L(S)$ to X. Moreover, $U_n(\phi)=\int_S x_n(s)\phi(s)d\alpha$ is for each n also defined and linear from $L(S)$ to X, and by Theorem 3.1.1 each U_n is weakly c.c. From (iii) and (iv) we also have $\lim_n |U-U_n| = \lim_n$ ess. $\sup_s \|x(s)-x_n(s)\| = 0$. Thus U is weakly c.c., by virt .e of the following fact: *the uniform limit of weakly c.c. operations is also weakly c.c.**

Conversely, if U is weakly c.c. and S is Euclidean, there is by Theorem 3.1.9 an $x(.) \in \mathfrak{A}^\infty(S)[X]$ such that $U(\phi)=\int_S x(s)\phi(s)d\alpha$ and $x(.)$ is essentially weakly compact-valued. Thus for some $E_0 \in \mathcal{E}_0$ the set $x(S-E_0)$ is separable and weakly compact in X. From an application of Lindelöf's theo-

* The proof of this statement is almost identical with that of the corresponding theorem for c.c. operations [1, Theorem 2, p. 96].

rem it follows [31, 1.12] that for each n there exists a countably-valued measurable function $x_n(s)$ defined over S, having $x_n(S-E_0) \subset x(S-E_0)$, and satisfying the inequality $\|x(s)-x_n(s)\| < 1/n$ over $S-E_0$. Each $x_n(.)$ is therefore in $\mathfrak{N}^\infty(S)[X]$ and is countably-valued and weakly compact-valued, and $\lim_n$ ess. $\sup._s \|x(s)-x_n(s)\| = 0$.

A similar criterion holds for complete continuity; namely, if either S is Euclidean or X is an $L^q(T)$ space, an operation U is c.c. from $L(S)$ to X if and only if essentially compact-valued $x_n(s)$ exist satisfying (i), (ii), and (iv). Furthermore, the $x_n(s)$ may always be chosen to be simple functions.

Part 2. Applications

In this part we take the opportunity to collect some of the preceding results, to translate them into kernel form, and then to apply them to obtain various conditions sufficient that a given kernel define a c.c. operation from $L(S)$ to X. Usually X will be taken to be an $L^q(T)$ space. A uniform mean ergodic theorem and an application to Markoff processes are discussed in sections D and E.

A. **Criteria for weak compactness in $L(S)$.** In order to apply the preceding results we shall require the following.

Theorem 3.2.1. *Let $[S, \mathcal{E}, \alpha]$ be any system. A subset Φ of $L(S)$ is weakly compact if and only if the following are satisfied:*

(1) *Φ is bounded,*

(2) *given $\epsilon > 0$ there is a $\delta > 0$ such that $\alpha(E) < \delta$ implies $\left| \int_E \phi(s) d\alpha \right| < \epsilon$ for every $\phi \in \Phi$,*

(3) *for each decomposition $\{S_i\}$ of S*

$$\lim_{m \to \infty} \sum_m^\infty \left| \int_{S_i} \phi(s) d\alpha \right| = 0$$

uniformly over Φ.*

If Φ is weakly compact, (1) is obvious. If (2) is false, there exist an $\epsilon > 0$, a sequence $\{\phi_n\}$ in Φ, and measurable sets $\{E_n\}$ such that

(4) $\alpha(E_n) < 1/n$ and $\left| \int_{E_n} \phi_n(s) d\alpha \right| > \epsilon$ *for* $n = 1, 2, \cdots$.

Since Φ is weakly compact, we may suppose $\{\phi_n\}$ to be weakly convergent. In particular, then,

(5) $\lim_{n \to \infty} \int_E \phi_n(s) d\alpha$ exists for every $E \in \mathcal{E}_B$.

But the space $\mathcal{E}_B$ is a complete metric space, and the argument based on the Baire category theorem that was used by Saks [35] for the case $\alpha(S) < \infty$ can be applied here to show that if (5) holds then the integrals $\int_E \phi_n(s) d\alpha$ are ab-

* In case $\alpha(T)$ is finite (3) is implied by (2).

solutely continuous over $\mathcal{E}_B$ uniformly with respect to n. This contradicts (4). Thus (2) holds. To see (3) let $\{S_i\}$ be any decomposition Δ of S. Then $U_\Delta(\phi) = \{\int_{S_i}\phi(s)d\alpha\}$ is an operation sending $L(S)$ into l. Since any operator with range in l takes weakly compact sets into compact sets, $U_\Delta(\Phi)$ is compact in l. Hence [4] (3) holds.

Before proving that (1), (2), and (3) imply that Φ is weakly compact we should like to note that $L(S)$ is weakly complete. For if $\{\phi_i\}$ converges weakly in $L(S)$ then $A_n(E) \equiv \int_E \phi_n(s)d\alpha$ is convergent for each $E \in \mathcal{E}$. Let $A(E) = \lim_n A_n(E)$. If $\{E_i\}$ are disjoint measurable sets then

$$A\left(\sum_{i\in\sigma} E_i\right) = \lim_n A_n\left(\sum_{i\in\sigma} E_i\right) = \lim_n \sum_{i\in\sigma} A_n(E_i)$$

for every set σ of positive integers. Hence if for each n the series $\sum_i A_n(E_i)$ is considered as an element ψ_n of l, it follows [1, p. 137] that $\{\psi_n\}$ is convergent in l. Thus there is a $\psi_0 \in l$ to which $\{\psi_n\}$ converges. The ith term in ψ_0 is the limit of the ith terms of the ψ_n's, that is, the ith term in the series ψ_0 is $\lim_n A_n(E_i) = A(E_i)$. Hence $\sum_i A(E_i)$ is absolutely convergent and $\sum_i A(E_i) = \lim_n \sum_i A_n(E_i) = A(\sum_i E_i)$, which means that $A(E)$ is completely additive over E. Since clearly $A(E) = 0$ when $\alpha(E) = 0$, $A(E)$ is the indefinite integral of some element ϕ_0 of $L(S)$. We thus have $\int_E \phi_0(s)d\alpha = \lim_n \int_E \phi_n(s)d\alpha$ for every E. Since $\{\phi_n\}$ converges weakly and finitely-valued functions are dense in $L^\infty(S)$, it then follows that ϕ_0 is the weak limit of the ϕ_n's.

Now suppose Φ satisfies (1), (2), and (3). Since $L(S)$ is weakly complete, it is enough to show that every sequence in Φ contains a subsequence that converges weakly. The finitely-valued functions being dense in $L^\infty(S)$ and Φ being bounded, it is sufficient to prove that in every sequence in Φ there is a subsequence $\{\phi_i\}$ such that $\lim_i \int_E \phi_i(s)d\alpha$ exists for every $E \in \mathcal{E}$. If $\alpha(S) < \infty$, it is known that (1) and (2) are necessary and sufficient that Φ be weakly compact [11]. Let $\{S_i'\}$ be a fixed decomposition of S. Given any sequence in Φ we can therefore find, since (1) and (2) are satisfied and $\alpha(S_i') < \infty$, a subsequence that is weakly convergent in $L(S_i')$. Thus by diagonalizing we may take the subsequence $\{\phi_n\}$ to have this property:

(6) $\lim_n \int_E \phi_n(s)d\alpha$ exists whenever $E \subset S_i'$ for some i.

Now let E in $\mathcal{E}$ be arbitrary but fixed. Set $S_{2i-1} = E \cdot S_i'$, $S_{2i} = S_i' - E \cdot S_i'$, so that $\{S_j\}$, $j = 1, 2, \cdots$, is a decomposition of S. From (3) there is for each $\epsilon > 0$ an N_ϵ such that

(7) $\sum_{j=2N_\epsilon+1} \left|\int_{S_j}\phi_n(s)d\alpha\right| < \epsilon$, $n = 1, 2, \cdots$.

Each S_j with $j \leqq 2N_\epsilon$ being a subset of some S_i', (6) implies that for some M_ϵ we have

(8) $\left|\int_{E_1}\phi_m(s)d\alpha - \int_{E_1}\phi_n(s)d\alpha\right| < \epsilon$ for $m, n \geqq M_\epsilon$,

where $E_1=\sum_{j=1}^{N_\epsilon}S_{2j-1}$. Writing E_2 for $\sum_{j=N^\epsilon+1}^{\infty}S_{2j-1}$, it follows from (8) and (7) that

$$\left|\int_E \phi_m-\int_E \phi_n\right| \leqq \sum_{i=1}^{2}\left|\int_{E_i}\phi_m-\int_{E_i}\phi_n\right|$$

$$<\epsilon+\sum_{j=N_\epsilon+1}^{\infty}\left|\int_{S_{2j-1}}\phi_m-\int_{S_{2j-1}}\phi_n\right|<2\epsilon$$

holds for $m, n\geqq M_\epsilon$. Thus $\lim_n\int_E\phi_n(s)d\alpha$ exists for each $E\in\mathcal{E}$, ending the proof.

THEOREM 3.2.2. *A set Φ in $L(S)$ is weakly compact if* (1) *and* (2) *are true and if*

(3′) $\lim_{m\to\infty}\sum_m^\infty\int_{S_i}|\phi(s)|d\alpha=0$ *uniformly over Φ for at least one decomposition $\{S_i\}$ of S.*

For (2) and (3′) imply (2) and (3).

B. **Kernel representations of operations among summable functions.** We prove first the following theorem.

THEOREM 3.2.3. *If for a fixed $q<\infty$ the real kernel $K(s, t)$ on $S\times T$ has the properties*

(i) *$K(s, t)$ is measurable on $S\times T$,*

(ii) *for almost all $s\in S$, $\int_T|K(s, t)|^q d\beta<\infty$,*

(iii) *for every $\psi'\in L^{q'}(T)$,*

$$\text{ess. sup.}_s\left|\int_T K(s, t)\psi'(t)d\beta\right|<\infty,$$

then

(iv) $\text{ess. sup.}_s\,(\int_T|K(s, t)|^q d\beta)^{1/q}\equiv C<\infty$,

(v) *the operation*

$$U(\phi)=\int_S K(s, t)\phi(s)d\alpha$$

is defined and separable from $L(S)$ to $L^q(T)$ and has the further properties

(vi) *U takes weakly compact sets into compact sets,**

(vii) *the norm $|U|$ of U is given by the constant C in* (iv),

(viii) *if $q>1$ then U is weakly c.c.*

For $q=1$ the statement (viii) *is not necessarily true but in this case if the kernel satisfies besides* (i), (ii), *and* (iii) *the condition*

(ix) *there is a null set E_0 such that* (a) $\lim_{\beta(F)=0}\int_F K(s, t)d\beta=0$ *uniformly on*

* See Theorem 3.2.7 and footnote.

$S-E_0$, (b) *for every decomposition* $\{T_i\}$ *of* T

$$\lim_{m\to\infty}\sum_{m}^{\infty}\left|\int_{T_i}K(s,t)d\beta\right|=0 \quad \textit{uniformly on } S-E_0,^*$$

then U *is weakly c.c. on* $L(S)$ *to* $L(T)$. *For* $1\leqq q<\infty$ *if the kernel* K *satisfies besides* (i), (ii), *and* (iii) *the condition*

(x) *for almost all* $s \in S$ *the set of points* $K(s, .)$ *in* $L^q(T)$ *is compact, then the operation* U *is c.c. from* $L(S)$ *to* $L^q(T)$.

Conversely, if U *is an arbitrary separable operation from* $L(S)$ *to* $L^q(T)$, $1<q<\infty$, *or if* S *is Euclidean and* α *is Lebesgue measure and* U *is a weakly c.c. operation from* $L(S)$ *to* $L(T)$, *there is a kernel* $K(s, t)$ *representing* U *and satisfying the conditions* (i)–(ix). *If* U *is c.c. from* $L(S)$ *to* $L^q(T)$, $1\leqq q\leqq\infty$, *there is a kernel* K *representing* U *and satisfying* (i)–(x).

If (i), (ii), and (iii) hold, it is seen from Theorem 1.3.6 that $x(s)=K(s, .)$ is in $\mathfrak{L}^\infty(S)[L^q(T)]$, that $U(\phi)=\int_S x(s)\phi(s)d\alpha=\int_S K(s, t)\phi(s)d\alpha$, and that (iv), (v), and (vii) are true. Since $U(\phi)=\int_S x(s)\phi(s)d\alpha$ the remaining conclusions result from Theorem 3.1.2 and the conditions for weak compactness in $L^q(T)$. The converse part of the theorem follows easily from Theorems 2.2.2, 3.1.7, 3.1.9, 3.1.10, and 1.3 5.

The next theorem, which has been given for Euclidean S by Dunford [9] and Gelfand [18], can be derived from the last theorem above on recalling conditions known to be necessary and sufficient for compactness in $L^q(T)$ when T is a linear interval [33, 40, 41].

THEOREM 3.2.4. *Let* T *be a real interval,* β *be Lebesgue measure, and* $1\leqq q<\infty$. *If* T *is bounded, then* U *is an operation defined and c.c. from* $L(S)$ *to* $L^q(T)$ *if and only if* $U(\phi)=\int_S K(s, t)\phi(s)d\alpha$ *where* $K(s, t)$ *satisfies* (i), (ii), *and* (iii) *of Theorem* 3.2.3 *and*

$$\lim_{h\to 0}\ \text{ess. sup.}_{s}\int_T|K(s,t+h)-K(s,t)|^q d\beta=0. \tag{3.2.41}$$

If T *is unbounded this statement remains true if* (3.2.41) *is supplemented by the condition that*

$$\lim_{N\to\infty}\ \text{ess. sup.}_{s}\int_T|K_N(s,t)-K(s,t)|^q d\beta=0$$

where $K_N(s, t)=K(s, t)$ *or* 0 *according as* $|t|\leqq N$ *or* $|t|>N$.

* In this part of the theorem, condition (ix)(b) can be replaced by the assumption that $\lim_{m\to\infty}\sum_m^\infty\int_{T_i}|K(s, t)|d\beta=0$ uniformly over $S-E_0$ for at least one decomposition $\{T_i\}$ of T: see Theorem 3.2.2.

As a second application of Theorem 3.2.3 (or of Theorem 3.2.4) there may be cited

THEOREM 3.2.5. *Let S be a real interval, T be $[0, 1]$, α and β be Lebesgue measure, and $K(r)$, $-\infty < r < \infty$, be periodic with period 1. Then if $K(.) \in L^q(T)$, $1 \leqq q < \infty$, the transformation*

$$U(\phi) = \psi(t) = \int_{-\infty}^{\infty} K(s-t)\phi(s)d\alpha$$

is defined and completely continuous from $L(S)$ to $L^q(T)$ and

$$|U| = \left(\int_0^1 |K(t)|^q d\beta\right)^{1/q}.$$

To prove this it is sufficient in view of Theorem 3.2.3 to verify the conditions (i), (iv), and (x) of that theorem. First we shall show that the kernel $K(s-t)$ is measurable. Using Lusin's theorem let E be a closed set on the interval $0 \leqq s \leqq 1$ such that $\alpha(E) > 1-\epsilon$ and $K(s)$ is continuous on E. Let E' be the set in $-1 \leqq s \leqq 0$ composed of all points of the form $s-1$ where $s \in E$. Let D be the set in the square $0 \leqq s, t \leqq 1$ consisting of those (s, t) for which $s-t \in E+E'$. Then D is a closed set and thus is measurable. To evaluate the measure of D we use Fubini's theorem on the characteristic function $\phi_D(s, t)$ of D. From the above construction it is clear that

$$\int_0^1 \phi_D(s, t)d\alpha = \alpha(E) > 1-\epsilon, \qquad 0 \leqq t \leqq 1,$$

and so $\alpha \times \beta(D) = \alpha(E) > 1-\epsilon$. Since $K(s-t)$ is continuous on D, it is, by Lusin's theorem, measurable over $[0, 1] \times [0, 1]$. From periodicity $K(s-t)$ is then measurable over $[n, n+1] \times [0, 1]$ for every integer n, and so (i) is satisfied. Periodicity also implies that

$$\operatorname*{ess.\,sup.}_s \int_T |K(s-t)|^q d\beta = \int_T |K(t)|^q d\beta = C < \infty,$$

which proves (iv). To prove (x) we use the known conditions for compactness in $L^q(T)$. Thus $K(s-.)$ is bounded and

$$\lim_{h \to 0} \int_T |K(s-t-h) - K(s-t)|^q d\beta = \lim_{h \to 0} \int_0^{-1} |K(u-h) - K(u)|^q du = 0$$

uniformly in s, so that (x) is satisfied.

The next theorem is analogous to Theorem 3.2.3.

THEOREM 3.2.6. *If the real kernel $K(s, t)$ on $S\times T$ has the properties*

(i) *$K(s, t)$ is measurable on $S\times T$,*

(ii) *there is a null set E_0 such that the set $K(s, .)$, $s \in S-E_0$, is a separable subset of $L^\infty(T)$, and*

(iii) *for every $\psi'(.) \in L(T)$*

$$\text{ess. sup.}_s \left| \int_T K(s, t)\psi'(t)d\beta \right| < \infty,$$

then

(iv) $\text{ess. sup.}_{s,t} |K(s, t)| \equiv C < \infty$,

(v) *the operation $U(\phi) = \int_S K(s, t)\phi(s)d\alpha$ is defined and separable from $L(S)$ to $L^\infty(T)$ and has the further properties*

(vi) *U takes weakly compact sets into compact sets,*

(vii) *the norm $|U|$ of U is given by the constant C in* (iv).

If the kernel $K(s, t)$ satisfies in addition to (i), (ii), *and* (iii) *the condition*

(viii) *for a null set E_0' the set $K(s, .)$, $s \in S-E_0'$, is weakly compact in $L^\infty(T)$,*

then U is weakly c.c. on $L(S)$ to $L^\infty(T)$. If the kernel $K(s, t)$ satisfies besides (i), (ii), *and* (iii) *the condition*

(ix) *for a null set E_0' the set $K(s, .)$, $s \in S-E_0'$, is compact in $L^\infty(T)$, then the operation U is c.c. from $L(S)$ to $L^\infty(T)$.*

Conversely, if S is a finite or infinite interval in Euclidean space and α is Lebesgue measure, and U is a weakly c.c. operator from $L(S)$ to $L^\infty(T)$, then there is a kernel $K(s, t)$ representing U and satisfying the conditions (i)–(viii). *If S is abstract and U is c.c., then* (ix) *is also satisfied.*

Properties (iv), (v), and (vii) come from Theorem 1.3.7. The remaining conclusions in the first part result from Theorem 3.1.2. The converse is derived by means of Theorems 3.1.9, 3.1.10, and 1.3.5.

The next result is essentially based on Theorems 3.1.2 and 3.2.1.

THEOREM 3.2.7. *Suppose $U(\phi) = \int_S K(s, t)\phi(s)d\alpha$ is an operation defined and separable from $L(S)$ to $L^q(T)$ and suppose $\Phi \subset L(S)$ satisfies conditions* (1)–(3) *of Theorem* 3.2.1. *Then the set*

$$\Psi = \left[\int_S K(s, t)\phi(s)d\alpha, \ \phi \in \Phi \right]$$

is compact in $L^q(T)$ when $1 < q < \infty$. This is also true when $q = 1$ provided that K is measurable and either* $\text{ess. sup.}_s \int_T |K(s, t)|\, d\beta < \infty$ *or S is Euclidean and T is linear.*

* Thus for example if $\alpha(E') < \infty$ and E varies over $\mathcal{E}(E')$ (see footnote to Theorem 2.1.0), the functions $\psi_E(t) = \int_E K(s, t)d\alpha$ form a compact set in $L^q(T)$.

Here in all three cases (Theorem 2.2.1, Theorem 2.4.5, and a remark at the end of the latter) U has a representation $\int_S x(s)\phi(s)d\alpha$ where $x(.) \in \mathfrak{N}^\infty(S)[L^q(T)]$. Hence U takes weakly compact sets into compact sets.

Another corollary to Theorem 3.2.3 is

THEOREM 3.2.8. *Suppose X is arbitrary. If $x(.) \in \mathfrak{N}^\infty(T)[X]$ and if $K(s, t)$ satisfies conditions* (i)–(iii) *and* (ix) *of Theorem* 3.2.3 *for $q=1$, then $U(\phi) = \int_T x(t)\{\int_S K(s, t)\phi(s)d\alpha\}d\beta$ is c.c. from $L(S)$ to X.*

For $\int_S K(s, t)\phi(s)d\alpha$ is weakly c.c. from $L(S)$ to $L(T)$ and $\int_T x(t)\psi(t)d\beta$ takes weakly compact sets in $L(T)$ into compact sets in X.

Of the several corollaries to the above theorem we mention only the following, obtained by letting $X = L^q(W)$.

THEOREM 3.2.9. *Suppose that $K(s, t)$ is given as in the preceding theorem. Let $[W, \mathcal{G}, \gamma]$ be a third system analogous to $[S, \mathcal{E}, \alpha]$ and $[T, \mathcal{F}, \beta]$. If $M(t, w)$ is measurable over $T \times W$, $M(t, .) \in L^q(W)$ for almost all t, and* ess. sup.$_t$ $|\int_W M(t, w)\chi(w)d\gamma| < \infty$ *for every $\chi(.) \in L^{q'}(W)$, then*

$$U(\phi) = \int_S \phi(s)\left\{\int_T K(s, t)M(t, w)d\beta\right\}d\alpha$$

is an operation defined and c.c. from $L(S)$ to $L^q(W)$, $q < \infty$. When $q = \infty$ this is still true provided that $M(t, .)$ is almost separably-valued in $L^q(W)$.

This can be shown to be a consequence of Theorem 3.2.8 and the Fubini theorem. A more direct way perhaps is as follows. We know that

$$L^q(W) \ni \int_T M(t, w)\psi(t)d\beta \qquad \text{for } \psi \in L(T),$$

and that in $\mathcal{E}_0$ there is an E_0 such that $K(s, .)$ forms a weakly compact set in $L(T)$ as s varies over $S - E_0$. Hence for each s in $S - E_0$

$$L^q(W) \ni \int_T M(t, w)K(s, t)d\beta \equiv H(s, w),$$

and $H(s, .)$, $s \in S - E_0$, forms a compact set in $L^q(W)$ since by Theorem 3.2.3 or 3.2.6 the operation $\int_T M(t, w)\psi(t)d\beta$ sends weakly compact sets into compact sets. Let $y(s) = H(s, .)$. From the measurability of M and K it follows that $H(s, w)$ is measurable, so that by Theorem 1.3.5 the vector function $y(.)$ is measurable. Since $y(.)$ is measurable and essentially compact-valued, the following operation is c.c. from $L(S)$ to $L^q(W)$:

$$\int_S y(s)\phi(s)d\alpha = \int_S H(s, w)\phi(s)d\alpha = \int_S \phi(s)\left\{\int_T K(s, t)M(t, w)d\beta\right\}d\alpha.$$

It is easily seen that a particular case of the last theorem is the following result of Servint [37; also see Theorems 1 and 3 of 46]: if $S=T=W=[0, 1]$ and all measures are that of Lebesgue, and if $K(s, t)$ and $M(t, w)$ are each bounded and measurable, then the above operation U is defined and c.c. from $L(S)$ to $L(W)$. In this case we can, by the above theorem, go even further and assert that U is defined and c.c. from $L(S)$ to $L^q(W)$ for each $q<\infty$, and that this is still true when $q=\infty$ provided that $M(t, .)$ is almost separably-valued in $L^\infty(W)$.

Translation of other theorems into kernel terminology will be left to the reader.

C. **C.c. operations from $L^p(S)$ to X.** We append the following remarks concerning c.c. operations from $L^p(S)$ to X. Here S is abstract and X is an arbitrary B-space.

THEOREM 3.2.10. *Let $x(.)$ on S to X be measurable and suppose that for some $p'<\infty$*

$$\|x(.)\| \equiv \left(\int_S \|x(s)\|^{p'} d\alpha\right)^{1/p} < \infty .$$

Then the equation

$$U(\phi) = \int_S x(s)\phi(s)d\alpha$$

defines a completely continuous operator from $L^p(S)$ to X with $|U| \leqq \|x(.)\|$.

The statement involving the norm follows from the Hölder inequality [9]. By the definition of measurability there is a sequence $x_n(s)$ of simple functions approaching $x(s)$ a.e. in S. Let $y_n(s)$ be defined to coincide with $x_n(s)$ on the set where $\|x_n(s)-x(s)\| \leqq \|x(s)\|$ and to vanish elsewhere. By the Lebesgue convergence theorem then

$$\lim_n \int_S \|x(s) - y_n(s)\|^{p'} d\alpha = 0.$$

Since the y_n's are simple they obviously define c.c. operators U_n which, by the above equation, satisfy the condition $\lim_n |U-U_n| = 0$. This shows that U is c.c.

THEOREM 3.2.11. *Let $K(s, t)$ be measurable on $S\times T$ and suppose that for some $p'<\infty$ and $q<\infty$*

$$\left(\int_S \left(\int_T |K(s, t)|^q d\beta\right)^{p'/q} d\alpha\right)^{1/p'} \equiv C < \infty .$$

Then the equation

$$\psi(t) = \int_S K(s, t)\phi(s)d\alpha$$

defines a completely continuous operator from $L^p(S)$ *to* $L^q(T)$.

This is an immediate corollary of the preceding theorem together with Theorems 1.3.6 and 1.3.5. This theorem was first proved for Euclidean S and T and $p=q$ by E. Hille and J. D. Tamarkin [21]. Another proof for Euclidean S and T was published later by F. Smithies [38].

As another application of Theorem 3.2.10 we mention the following.

THEOREM 3.2.12. *Let* $K(s, t)$ *be measurable on* $S\times T$ *and such that the function* $K(s, .)$ *is almost separably-valued in* $L^\infty(T)$. *Then if* $p' < \infty$ *and*

$$\int_S \operatorname{ess.\,sup.}_t |K(s, t)|^{p'} d\alpha < \infty,$$

the operation

$$U(\phi) = \int_S K(s, t)\phi(s)d\alpha$$

is defined and c.c. from $L^p(S)$ *to* $L^\infty(T)$.

This is immediately derivable from Theorems 1.3.6, 1.3.5, and 3.2.10.

Further applications of Theorem 3.2.10 will be left to the reader.

D. **A mean ergodic theorem.** Let U be an operation sending $L(S)$ into $L(S)$ and having a representation $U(\phi)=\int_S x(s)\phi(s)d\alpha$ where $x(.)$ is an essentially weakly compact-valued element of $\mathfrak{A}^\infty(S)[L(S)]$. Since U is separable, $U(L(S))$ has a separable span X in $L(S)$. From the footnote to Theorem 2.2.6 there is a Borel field $\mathcal{F}$ contained in $\mathcal{E}$ such that $\mathcal{F}_B$ is separable and X is a subset of those elements of $L(S)$ which are $\mathcal{F}$-measurable. If $\mathcal{F}$ is any such Borel field and if we set $T=S$ and $\beta=\alpha$, the triad $[T, \mathcal{F}, \beta]$ is a system for which $L(T)$ is a separable c.l.m. within the $\mathcal{F}$-measurable elements of $L(S)$, and $L(T)\supset X\supset U(L(S))$.

The particular mean ergodic theorem we wish to give is one dealing with operations of this sort.* Basically it is a special case of an extension made to general B-spaces by Fortet [14, 13], Yosida [43, 44], and Kakutani [22] of a result due to Kryloff and Bogoliouboff [26] and Fréchet [15].

* In *Dependent probabilities and spaces* (*L*), Proceedings of the National Academy of Sciences, vol. 24 (1938), pp. 154–159, Garrett Birkhoff has discussed for generalized $L(S)$ spaces the discrete mean ergodic theory of what he has termed "transition operators" (also see S. Kakutani, *Mean ergodic theorem in abstract* (*L*)*-spaces*, Proceedings of the Imperial Academy, Tokyo, vol. 15 (1939), pp. 121–123). These are norm-preserving operators that send non-negative elements into non-negative elements. Some account of the continuous theory in an $L(S)$ space can be found in [11].

THEOREM 3.2.13. *If U is an operation of the above type, these conclusions can be made*: (I) *there is a kernel $K_1(s, t)$ defined over $S\times T = S\times S$ such that*

(i) *$K_1(s, t)$ is measurable,*

(ii) *$K_1(s, .) = x(s)$ in $L(T)$ for almost every s, and*

(iii) *$U(\phi) = \int_S K_1(s, t)\phi(s)d\alpha$ for each $\phi \in L(S)$;*

(II) (i) *U is separable and weakly c.c.,*

(ii) *U takes weakly compact sets into compact sets,*

(iii) *U^n is c.c. for $n \geqq 2$,*

(iv) *$|U| = \text{ess. sup.} \|x(s)\| = \text{ess. sup.}_s \int_T |K_1(s, t)|\, d\beta$,*

(v) *the maximum number of linearly independent solutions of the equation $\phi = U(\phi)$ is a finite number N.*

Suppose in addition that the norms $|U^n|$ are bounded. Then

(vi) *$\lim_{m,n} |V_m - V_n| = 0$ where $V_m = (1/m)\sum_1^m U^i$.*

(III) *The limit operation $V = \lim V_m$ has these properties*:

(i) *V is c.c.,*

(ii) *$VU^n = U^nV = V$ for $n = 1, 2, \cdots$,*

(iii) *$V^2 = V$,*

(iv) *the fixed points of V form a c.l.m. Y having dimension N,*

(v) *Y is exactly the set of fixed points of U,*

(vi) *$V(L(S)) = Y$.*

(IV) *If $x_0(s) \equiv V(x(s))$ then*

(i) *$x_0(.) \in \mathfrak{A}^\infty(S)[Y]$,*

(ii) *$V(\phi) = \int_S x_0(s)\phi(s)d\alpha$,*

(iii) *$|V| = \text{ess. sup.} \|x_0(s)\|$,*

(iv) *$x_0(s) = V(x(s)) = U^n(x_0(s)) = V(U^n(x(s))) = V(x_0(s))$ for every n and almost all s,*

(v) *$x_0(s)$ can be written as $x_0(s) = \sum_1^N \mu_i(s)\phi_i$ where*

(vi) *each ϕ_i is in Y and $\|\phi_i\| = 1$,*

(vii) *the set $\phi_1, \cdots, \phi_N$ forms a basis for Y,*

(viii) *$\mu_i(.) \in L^\infty(S)$ and $\int_S \mu_i(s)\phi_i(s) = 1$ or equals 0 according as $i = j$ or $i \neq j$, and*

(ix) *in $L^\infty(S)$ the c.l.m. Y^* generated by the set $\mu_i(.)$, $i = 1, 2, \cdots, N$, is both the set of fixed points of U^* (the operation adjoint to U) and the set of fixed points for the c.c. adjoint V^* of V, and Y^* has dimension N.*

(V) *Finally, the following statements are true concerning the measurable kernels $K_0(s, t) \equiv \sum_1^N \mu_i(s)\phi_i(t)$, $K_1(s, t)$, and $K_n(s, t) = \int_S K_{n-1}(s, s')K_1(s', t)d\alpha$, $n = 2, 3, \cdots$:*

(i) *$V(\phi) = \int_S K_0(s, t)\phi(s)d\alpha$,*

(ii) *$U^n(\phi) = \int_S K_n(s, t)\phi(s)\, d\alpha$ for $n = 1, 2, \cdots$,*

(iii) *for almost every s the equality*

$$K_0(s, t) = \int_S K_{n+1}(s, s')K_0(s', t)d\alpha = \int_S K_0(s, s')K_n(s', t)d\alpha$$

holds for $n=0, 1, 2, \cdots$ *and for almost all* t;

(iv) $K_n(s, .)$ *is essentially weakly compact-valued in* $L(T)$ *for* $n=1$ *and essentially compact-valued for* $n\neq 1$.

There is a null set E_0 *such that*

(v) $\lim_{n\to\infty} \int_T |K_0(s, t)-(1/n)\sum_1^n K_i(s, t)|\, d\beta=0$ *uniformly over* $S-E_0$,

(vi) $\lim_{n\to\infty}(1/n)\sum_1^n \int_F K_i(s, t)d\beta=\int_F K_0(s, t)d\beta$ *uniformly over* $(S-E_0)\times F$,

(vii) $\lim_{n\to\infty} \int_T \left|\int_S \phi(s)K_0(s, t)d\alpha-\int_S(\phi(s)/n)\sum_1^n K_i(s, t)d\alpha\right| d\beta=0$ *uniformly over those* ϕ *in* $L(S)$ *having* $\int_S|\phi(s)|\, d\alpha\leqq 1$.

Statements (I)(i)–(iii) and (II)(iv) are contained in Theorem 1.3.5, and Theorem 3.1.6 yields (II)(i)–(iii) and (II)(v). If the norms $|U^n|$ are bounded, conclusions (II)(vi) and (III)(i)–(iii) follow from [43] or Theorem 5 of [22] and the fact that U^2 is c.c. The definition of V taken together with (III)(ii) shows that the fixed points of V, which form a c.l.m. Y, are also the fixed points of U; this vindicates (III)(iv)–(v). Conclusion (III)(vi) comes from (III)(iii) and the definition of Y. Letting $x_0(s)\equiv V(x(s))$, (III)(vi) shows that $x_0(S)\subset Y$; and since measurability and being essentially bounded are two properties that are operational invariants, (IV)(i) is true. To verify (IV)(ii) it is sufficient to note that $V(\phi)=V(U(\phi))=V(\int_S x(s)\phi(s)d\alpha)=\int_S V(x(s))\phi(s)d\alpha$; (IV)(iii) is then obvious. Conclusion (IV)(iv) results easily from (III)(ii)–(iii), the definition of $x_0(s)$, and the assumption that $x(s)$ is defined for almost every s.

To establish the rest of (IV) we may proceed as follows. Since Y is a c.l.m. in $L(S)$ having finite dimension N, it is possible to choose in Y a basis $\phi_1, \cdots, \phi_N$ satisfying (IV)(vi). Choosing such a basis we have

$$\phi = \sum_1^N g_i(\phi)\phi_i, \qquad \phi \in Y,$$

where each g_i is a linear functional over Y and $g_i(\phi_j)=\delta_{ij}$, the Kronecker delta. Thus

$$x_0(s) = \sum_1^N g_i(x_0(s))\phi_i = \sum_1^N \mu_i(s)\phi_i$$

where $\mu_i(.)\equiv g_i(x_0(.))$ is in $L^\infty(S)$ for each i since $x_0(.)\in \mathfrak{A}^\infty(S)[Y]$. Statements (IV)(v)–(viii) are therefore verified. To establish (IV)(ix) we first note that since $\lim_m |V-V_m|=0$ we also have $\lim_m |V^*-V_m^*|=0$ where V^* and V_m^* are the operations adjoint to V and V_m respectively and $V_m^*=(1/m)\sum_1^m (U^i)^*=(1/m)\sum_1^m (U^*)^i$. It then follows that U^* and V^* have

the same set of fixed points, and this set is a c.l.m. in $L^\infty(S)$ that has dimension N since V^* is the adjoint of the c.c. operation V. This c.l.m. is Y^*. For (IV)(ii) and Theorem 1.1.3 imply that for any f in the adjoint of $L(T)$ the linear functional $V^*(f)$ over $L(S)$ is defined by the element $f(x_0(.))$ of $L^\infty(S)$. Since $f(x_0(s)) = \sum_1^N \mu_i(s) f(\phi_i)$, $V^*(f)$ is defined by an element of Y^*. Thus $V^*(L^\infty(T)) \subset Y^*$, and hence $f \in Y^*$ when $V^*(f) = f$. Conversely, if $f \in Y^*$ then $V^*(f) = f$. It is sufficient to consider the functional f_j defined by the element $\mu_j(.)$ of $L^\infty(T)$. For each ϕ_i we have

$$f_j(\phi_i) = \int_S \mu_j(s)\phi_i(s)d\alpha = \int_S g_j(x_0(s))\phi_i(s)d\alpha = g_j(V(\phi_i)) = g_j(\phi_i),$$

so that $f_j(x_0(s)) = \sum_1^N f_j(\phi_i)\mu_i(s) = \sum_1^N g_j(\phi_i)\mu_i(s) = \mu_j(s)$. Thus $V^*(f_j)$ is defined by $\mu_j(s)$, that is, $V^*(f_j) = f_j$. This completes the proof of (IV)(ix).

In (V) the first conclusion results from (IV)(ii) and (IV)(v). Setting

(1) $x_1(s) \equiv x(s)$ and $x_n(s) = U(x_{n-1}(s))$ for $n \geqq 2$,

it is then clear by induction that the three statements

(2) $x_n(.) \in \mathfrak{A}^\infty(S)[L(S)]$,

(3) $U^n(\phi) = \int_S x_n(s)\phi(s)d\alpha$,

(4) $K_n(s, .) = x_n(s)$ in $L(S)$ for almost all s

must each hold for $n = 1, 2, \cdots$. Since each $K_n(s, t)$ is measurable, (V)(ii) then follows from (2), (3), (4), and Theorem 1.3.5. Conclusion (V)(iii) comes from (1), (3), (4), and (V)(ii), and (V)(iv) is a corollary to (2), (3), (4), (II)(i), (II)(iii), and Theorem 3.1.2. Finally, (V)(v) and (V)(vii) are true since

$$0 = \lim_m |V - V_m| = \lim_m \text{ess. sup.} \left\| x_0(s) - \frac{1}{m}\sum_1^m x_i(s) \right\|$$

$$= \lim_m \text{ess. sup.}_s \int_T \left| K_0(s, t) - \frac{1}{m}\sum_1^m K_i(s, t) \right| d\beta,$$

and (V)(vi) is obvious from (V)(v).

This and Theorem 3.1.9 lead at once to

THEOREM 3.2.14. *If S is a finite or infinite Euclidean interval and α is Lebesgue measure and U is an arbitrary weakly c.c. operation in $L(S)$, then $U(\phi) = \int_S x(s)\phi(s)d\alpha$ where $x(.)$ is an essentially weakly compact-valued element of $\mathfrak{A}^\infty(S)[L(S)]$. Hence conclusions* (I) *and* (II)(i)–(v) *of Theorem* 3.2.13 *hold in this case, and if the norms $|U^n|$ are bounded the remaining conclusions are also valid for U.*

A simple example shows that in Theorem 3.2.13 we cannot drop entirely the assumption that $x(.)$ is essentially weakly compact-valued. On the real

axis let J_n be the interval $[n, n+1)$ for each integer n and on each J_n assign $x(s)$ the constant value ϕ_{n+1} where ϕ_{n+1} is the characteristic function of J_{n+1}. Obviously $x(.) \in \mathfrak{A}^\infty(S)[L(S)]$ and ess. sup. $\|x(s)\| = 1$, so that $U(\phi) = \int_S x(s)\phi(s)ds$ satisfies all the hypotheses of Theorem 3.2.13 except that $x(.)$ is not essentially weakly compact-valued. Since $U(\phi_n) = \phi_{n+1}$, it follows that the principal conclusions of the theorem are false in this case. For $\Phi = \{\phi_n\}$ is bounded in $L(S)$ but not weakly compact, and thus U^n fails for each n to be weakly c.c. since $U(\Phi) = \Phi$. Moreover, $\|V_m(\phi_0) - V_{2m}(\phi_0)\| = 1$ for every m, as the reader can easily verify; hence the operations V_m do not converge even pointwise.

E. **An application to Markoff processes.** Let $\mathcal{J}$ be a second Borel field of subsets of S, with $\mathcal{J} \subset \mathcal{E}$ and $\mathcal{J}_B$ separable. Take $T = S$ and $\beta = \alpha$, so that $[T, \mathcal{J}, \beta]$ forms a second system for which $L(T)$ is a separable c.l.m. in $L(S)$. Suppose further that there is a function $P(s, F)$ defined over $S \times \mathcal{J}$ and having these properties:

(1) $P(s, F)$ is $\mathcal{E}$-measurable in s for each $F \in \mathcal{J}_B$,

(2) $s \in S$ implies that $P(s, F)$ is non-negative and completely additive over $\mathcal{J}$ and $P(s, T) = 1$,

(3) if $F_1 \supset F_2 \supset \cdots \supset F_n \supset \cdots$ and $\beta(\prod_1^\infty F_n) = 0$, then $\lim_n P(s, F_n) = 0$ uniformly over S.

As Doob has shown [6], under these circumstances $P(s, F)$ is a set of conditional probability functions for a certain class of Markoff processes. Functions $P(s, F)$ of this sort have been discussed frequently in probability (for references see [6]), and most of the statements in the next theorem are already known, being due either to Fréchet [15], Kryloff and Bogoliouboff [26], Doeblin and Fortet [5], Doob [6], Yosida and Kakutani [45], or to earlier authors. However, the indigeneity of the theorem and the fact that some of the conclusions are apparently new lead us to include it here.

From (2), (3), and the Radon-Nikodym theorem there is for each s a point $x(s)$ in $L(T)$ whose indefinite integral over F is precisely $P(s, F)$. We now make the assertion that

(4) *$x(.)$ is a weakly compact-valued element of $\mathfrak{A}^\infty(S)[L(T)]$, $\|x(s)\| \equiv 1$, and each point $x(s)$ is an essentially non-negative element of $L(T)$.*

In the first place it is evident from (2) that for each s the element $x(s)$ in $L(T)$ is essentially non-negative and has $\|x(s)\| = 1$. Secondly, in view of (1), the decomposability of T, and the complete additivity in F of $P(s, F)$, it results that $P(s, F)$ is measurable in s for each $F \in \mathcal{J}$. Hence $f(x(s))$ is measurable for each linear functional f defined over $L(T)$ by a finitely-valued element of $L^\infty(T)$. Since such elements are dense in $L^\infty(T)$, $f(x(s))$ must be measurable for every linear functional over $L(T)$. The space $L(T)$ being separable, it then

follows from Theorem 1.1.7 that $x(.)$ ε $\mathfrak{A}^{\infty}(S)[L(T)]$ and $\|x(s)\| \equiv 1$. To see that $x(.)$ is weakly compact-valued we have only to note that property (3) of $P(s, F)$ implies that $\Phi = x(S)$ satisfies conditions (2) and (3) of Theorem 3.2.1.

Since $L(T)$ forms a c.l.m. in $L(S)$ and $L(T) \supset x(S)$, we can also state that

(5) *assertion* (4) *above holds with* $L(T)$ *replaced by* $L(S)$.

We are now ready to give the theorem. Here, as before, $\mathfrak{P}$ denotes the essentially non-negative elements of $L(S)$ and $\mathfrak{P}_1 = \mathfrak{P}[\|\phi\| = 1]$.

THEOREM 3.2.15. *Let* U *be the operation* $U(\phi) = \int_S x(s)\phi(s)d\alpha$ *defined by the function* $x(.)$ *of* (4). *Then all the conclusions of Theorem* 3.2.13 *hold here and in addition*

(I) (iv) $\int_F K_1(s, t)d\beta = P(s, F)$ *for every* s ε S *and every* F ε $\mathfrak{F}$,

(v) $K_1(s, t) \geqq 0$,

(vi) $\int_T K_1(s, t)d\beta = 1$ *for every* s,

(II) (vii) $|U| = 1$,

(viii) *if* ϕ ε $\mathfrak{P}$ *then* $U(\phi)$ ε $\mathfrak{P}$ *and* $\|U(\phi)\| = \|\phi\|$,

(III) (vii) $|V| = 1$,

(viii) *if* ϕ ε $\mathfrak{P}$ *then* $V(\phi)$ ε $\mathfrak{P}$ *and* $\|V(\phi)\| = \|\phi\|$,

(ix) *the dimension number* N *of* Y *is positive*,

(IV) (x) *each* ϕ_i *in the basis* $\phi_1, \cdots, \phi_N$ *may be chosen in* $\mathfrak{P}_1$,

(V) (viii) *for* $n = 0, 1, 2, \cdots$, $K_n(s, t) \geqq 0$ *a.e. in* $S \times T$ *and* $\int_T K_n(s, t)d\beta = 1$ *a.e. in* S.

From (5) and the equality $|U| = \text{ess. sup. } \|x(s)\| = 1$ it is evident that U satisfies the requirements of Theorem 3.2.13, including the one that the norms $|U^n|$ be bounded. Moreover, the span of $U(L(S))$ must by Theorem 1.2.10 be in the span X belonging to $x(S)$; thus $U(L(S)) \subset X \subset L(T)$, and X is clearly a subset of those elements of $L(S)$ which are $\mathfrak{F}$-measurable. Hence all the conclusions of Theorem 3.2.13 hold in the present situation.

The kernel K_1 given in Theorem 3.2.13 can fail to satisfy (I)(iv) and (I)(v) on at most a set of $S \times T$ measure zero. We may suppose K_1 altered on this set so as to satisfy these two conditions; the previous properties of K_1 and of the kernels K_n will not be affected since the exceptional set was of measure 0. Conclusion (II)(vii) was shown above, and (I)(vi) is obvious from (I)(iv). According to (4) the closed convex set $\mathfrak{P}_1$ contains $x(S)$; from (II) of Theorem 1.2.10 we conclude that

(6) $U(\phi)$ ε $\mathfrak{P}_1$ for each ϕ ε $\mathfrak{P}_1$.

This yields (II)(viii). On writing $x_1(s) \equiv x(s)$ and $x_i(s) = U(x_{i-1}(s))$ for $i \geqq 2$ we noted, in proving Theorem 3.2.13, that

(7) $U^i(\phi) = \int_S x_i(s)\phi(s)d\alpha$.

From (6) it follows that $x_i(s) \in \mathfrak{P}_1$ for each $i \geqq 1$ and each s, so that $(1/n)\sum_1^n x_i(s)$ is in the convex set $\mathfrak{P}_1$ for every n and s. Now (7) implies that

$$\lim_n \text{ess. sup.}_s \left\| x_0(s) - \frac{1}{n}\sum_1^n x_i(s) \right\| = \lim_n |V - V_n| = 0,$$

and hence $x_0(s) = \lim_n (1/n)\sum_1^n x_i(s)$ for almost every s. Since $\mathfrak{P}_1$ is closed we must have $x_0(s) \in \mathfrak{P}_1$ for almost all s. This combined with Theorem 1.2.10 shows that (III)(viii) holds and that $|V| = \text{ess. sup.} \|x_0(s)\| = 1$. Conclusion (III)(ix) now follows from the fact that $Y = V(L(S))$ and that $|V| \neq 0$. To prove (IV)(x) let $\phi_1', \cdots, \phi_N'$ be any basis for Y with $\|\phi_i'\| = 1$. From (III)(iv), (vi), and (vii) the unit sphere Y_1 of Y is the image under V of the unit sphere of $L(S)$. Theorem 1.2.10 then implies that for each null set E_0 the set Y_1 is in the closed convex hull of the point set sum $Y_2 = x_0(S-E_0) + R(x_0(S-E_0))$. Choosing E_0 such that $x_0(s) \in \mathfrak{P}_1$ for $s \notin E_0$, it follows that the closed convex hull of $x_0(S-E_0)$ is the convex hull of $x_0(S-E_0)$. The same is therefore true of Y_2. Hence each $\phi \in Y_1$ can be written as $\phi = \sum_{j=1}^k c_j \bar{\phi}_j$ where each $\bar{\phi}_j \in \mathfrak{P}_1$ and $\sum_1^k |c_j| = 1$. On expressing each ϕ_i' as $\phi_i' = \sum_{j=1}^{k_i} c_{ij}\bar{\phi}_{ij}$ it is clear that each element of Y is a linear combination of the $\bar{\phi}_{ij}$'s. Hence from among these a subset $\phi_1, \cdots, \phi_N$ can be chosen to form a basis for Y. This establishes (IV)(x). Finally, (V)(viii) results on recalling that $x_n(s) = K_n(s, .)$ in $L(T)$ for almost all s and that $x_n(s) \in \mathfrak{P}_1$ for every s.

References

1. S. Banach, *Théorie des Opérations Linéaires*, Warsaw, 1932.
2. G. Birkhoff, *Integration of functions with values in a Banach space*, these Transactions, vol. 38 (1935), pp. 357–378.
3. S. Bochner, *Integration von Funktionen, deren Werte die Elemente eines Vektorraumes sind*, Fundamenta Mathematicae, vol. 20 (1933), pp. 262–276.
4. L. W. Cohen and N. Dunford, *Linear operations on sequence spaces*, Duke Mathematical Journal, vol. 3 (1937), pp. 689–701.
5. W. Doeblin and R. Fortet, *Sur deux notes de MM. Kryloff et Bogoliouboff*, Comptes Rendus de l'Académie des Sciences, Paris, vol. 204 (1937), pp. 1699–1701.
6. J. L. Doob, *Stochastic processes with integral-valued parameter*, these Transactions, vol. 44 (1938), pp. 87–150.
7. ———, *One-parameter families of transformations*, Duke Mathematical Journal, vol. 4 (1938), pp. 752–774.
8. N. Dunford, *Integration in general analysis*, these Transactions, vol. 37 (1935), pp. 441–453; corrections, ibid., vol. 38 (1936), pp. 600–601.
9. ———, *Integration and linear operations*, these Transactions, vol. 40 (1936), pp. 474–494.
10. ———, *Uniformity in linear spaces*, these Transactions, vol. 44 (1938), pp. 305–356.
11. ———, *A mean ergodic theorem*, Duke Mathematical Journal, vol. 5 (1939), pp. 635–646.
12. M. Eidelheit, *Zur Theorie der Konvexen Mengen in linearen normierten Räumen*, Studia Mathematica, vol. 6 (1936), pp. 104–111.
13. R. Fortet, *Sur des probabilités en chaîne*, Comptes Rendus de l'Académie des Sciences, Paris, vol. 204 (1937), pp. 315–317.

14. ———, *Sur des opérations linéaires non complètement continus*, Comptes Rendus de l'Académie des Sciences, Paris, vol. 204 (1937), pp. 1543–1545.

15. M. Fréchet, *Sur l'allure asymptotique de la suite des itérés d'un noyau de Fredholm*, Quarterly Journal of Mathematics, vol. 5 (1934), pp. 106–144.

16. V. Gantmakher and V. Smulian, *Sur les espaces dont la sphère unitaire est faiblement compacte*, Comptes Rendus de l'Académie des Sciences de l'URSS, vol. 17 (1937), pp. 91–94.

17. I. Gelfand, *Sur un lemme de la théorie des espaces linéaires*, Communications de l'Institut des Sciences Mathématiques et Mécaniques de l'Université de Kharkoff, (4), vol. 13 (1935), pp. 35–40.

18. ———, *Abstrakte Funktionen und lineare Operatoren*, Recueil Mathématique, new series, vol. 4 (1938), pp. 235–284.

19. T. H. Hildebrandt, *Lebesgue integration in general analysis*, Bulletin of the American Mathematical Society, vol. 33 (1927), p. 646.

20. ———, *Linear operations on functions of bounded variation*, Bulletin of the American Mathematical Society, vol. 44 (1938), pp. 75–76.

21. E. Hille and J. D. Tamarkin, *On the theory of linear integral equations* II, Annals of Mathematics, (2), vol. 35 (1934), pp. 445–455.

22. S. Kakutani, *Iteration of linear operations in complex Banach spaces*, Proceedings of the Imperial Academy, Tokyo, vol. 14 (1938), pp. 295–300.

23. L. Kantorovitch, *Allgemeine Formen gewisser Klassen von linearen Operatoren*, Comptes Rendus de l'Académie des Sciences de l'URSS, vol. 12 (1936), pp. 101–106.

24. L. Kantorovitch and B. Vulich, *Sur la représentation des opérations linéaires*, Compositio Mathematica, vol. 5 (1937), pp. 119–165.

25. ———, *Sur un théorème de M. N. Dunford*, ibid., vol. 5 (1938), pp. 430–432.

26. N. Kryloff and N. Bogoliouboff, *Sur les probabilités en chaîne*, Comptes Rendus de l'Académie des Sciences, Paris, vol. 204 (1937), pp. 1386–1388.

27. S. Mazur, *Über konvexen Mengen in linearen normierten Räumen*, Studia Mathematica, vol. 4 (1933), pp. 70–84.

28. ———, *Über die kleinste konvexe Menge, die eine gegebene kompakte Menge enthält*, Studia Mathematica, vol. 2 (1930), pp. 7–9.

29. O. Nikodym, *Sur une généralisation des intégrales de M. J. Radon*, Fundamenta Mathematicae, vol. 15 (1930), pp. 131–179.

30. ———, *Contribution à la théorie des fonctionelles linéaires* · · · , Mathematica, vol. 5 (1931), pp. 130–141.

31. B. J. Pettis, *On integration in vector spaces*, these Transactions, vol. 44 (1938), pp. 277–304.

32. ———, *Differentiation in Banach spaces*, Duke Mathematical Journal, vol. 5 (1939), pp. 254–269.

33. M. Riesz, *Sur les ensembles compacts de fonctions sommables*, Acta Szeged, vol. 6 (1933), pp. 136–142.

34. S. Saks, *Theory of the Integral*, Warsaw and Lwów, 1937.

35. ———, *Addition to the note on some functionals*, these Transactions, vol. 35 (1933), pp. 965–970.

36. W. Sierpiński, *Sur un problème concernant les ensembles measurables superficiellement*, Fundamenta Mathematicae, vol. 1 (1920), pp. 112–115.

37. J. Servint, *Sur les transformations intégrales de l'espace L*, Comptes Rendus de l'Académie des Sciences de l'URSS, vol. 18 (1938), pp. 255–257.

38. F. Smithies, *A note on completely continuous transformations*, Annals of Mathematics, (2), vol. 38 (1937), pp. 626–630.

39. W. L. Chmoulyan, *Sur les ensembles faiblement compacts dans les espaces linéaires normés*, Communications de l'Institut des Sciences Mathématiques et Mécaniques de l'Université de Kharkoff, (4), vol. 14 (1937), pp. 239–242.

40. J. D. Tamarkin, *On the compactness of the space L_p*, Bulletin of the American Mathematical Society, vol. 38 (1932), pp. 79–84.

41. A. Tulajkov, *Zur Kompaktheit im Raum L_p für $p=1$*, Nachrichten der Gesellschaft der Wissenschaften, Göttingen, 1933, pp. 169–170.

42. B. Vulich, *Sur les formes générales de certaines opérations linéaires*, Recueil Mathematiques, new series, vol. 2 (1937), pp. 275–305.

43. K. Yosida, *Abstract integral equations and the homogeneous stochastic process*, Proceedings of the Imperial Academy, Tokyo, vol. 14 (1938), pp. 286–291.

44. ———, *Mean ergodic theorem in Banach spaces*, ibid., pp. 292–294.

45. K. Yosida and S. Kakutani, *Application of the mean ergodic theorem to the problems of Markoff's process*, ibid., pp. 333–339.

46. K. Yosida, Y. Mimura, and S. Kakutani, *Integral operator with bounded kernel*, ibid., pp. 359–362.

YALE UNIVERSITY,
NEW HAVEN, CONN

NILPOTENT GROUPS AND THEIR GENERALIZATIONS*

BY
REINHOLD BAER

Nilpotent finite groups may be defined by a great number of properties. Of these the following three may be mentioned, since they will play an important part in this investigation. (1) The group is swept out by its ascending central chain (equals its hypercentral). (2) The group is a direct product of p-groups (that is, of its primary components). (3) If S and T are any two subgroups of the group such that T is a subgroup of S and such that there does not exist a subgroup between S and T which is different from both S and T, then T is a normal subgroup of S. These three conditions are equivalent for finite groups; but in general the situation is rather different, since there exists a countable (infinite) group with the following properties: all its elements not equal to 1 are of order a prime number p; it satisfies condition (3); its commutator subgroup is abelian; its central consists of the identity only.

A group may be termed soluble, if it may be swept out by an ascending (finite or transfinite) chain of normal subgroups such that the quotient groups of its consecutive terms are abelian groups of finite rank. A group satisfies condition (1) if, and only if, it is soluble and satisfies condition (3) (§2); and a group without elements of infinite order satisfies (1) if, and only if, it is the direct product of soluble p-groups (§3); and these results contain the equivalence of (1), (2) and (3) for finite groups as a trivial special case. If a group without elements of infinite order may be swept out by an ascending chain of subgroups such that each is a normal subgroup of the next one and such that the quotient groups of its consecutive terms are cyclic, then (2) and (3) are equivalent properties, though they no longer imply (1) (§4). If a group satisfies condition (1)—or suitable weaker conditions—then the elements of finite order in this group generate a subgroup without elements of infinite order which is a direct product of p-groups.

A seemingly only slightly stronger condition than (3) is the following property: (3′) If S and T are any two subgroups of the group such that T is a subgroup of S and such that there exists at most one subgroup between S and T which is different from both S and T, then T is a normal subgroup of S. Clearly (3′) implies (3), though there exist groups which satisfy (3), but not (3′). A closer investigation reveals however that (3′) is a much stronger imposition than it seems to be, since it is possible to prove the following theorem: A group, that either does not contain elements of infinite

* Presented to the Society, September 5, 1939; received by the editors August 28, 1939.

order or else is swept out by its ascending central chain, satisfies condition (3′) if, and only if, all its subgroups are normal subgroups, that is, if it is either abelian or hamiltonian (§6).

In an appendix we have given a list of properties which may serve as a definition of nilpotent groups, provided the group in question is finite; and since these properties are in general no longer equivalent, we have added a chart, indicating their interrelations.

0. In this section we state some notations and facts which will be used in the course of this investigation.

0.A. $Z(G)$ is the *central* of the group G.

$Z_0(G)=1$; $Z_\nu(G)\leqq Z_{\nu+1}(G)\leqq G$ and $Z(G/Z_\nu(G))=Z_{\nu+1}(G)/Z_\nu(G)$; if λ is a limit-ordinal, then $Z_\lambda(G)$ is the join of all the $Z_\nu(G)$ for $\nu<\lambda$, so that the ascending central chain $Z_\nu(G)$ is defined for every finite or infinite ordinal ν.

$C(G)$ is the *commutator subgroup* of G.

$C_0(G)=G$, $C_{i+1}(G)=C(C_i(G))$ so that $C_i(G)$ is defined for every integer i.

$C^0(G)=G$, $C^{i+1}(G)$ is the subgroup generated by all the elements $gcg^{-1}c^{-1}$ for g in G, c in $C^i(G)$, so that the descending chain $C^i(G)$ is defined for every integer i.

0.B. *Every subgroup of the group G is a normal subgroup of G* if, and only if, $x^{-1}yx$ is a power of y for every pair x, y of elements in G. If every subgroup of G is a normal subgroup of G, then there are two possibilities:

(A) G is abelian, that is, $C(G)=1$, $Z(G)=G$.

(H) G is hamiltonian. Then $C(G)$ is of order 2, $Z(G)$ consists of those elements whose order is not divisible by 4, $G/Z(G)$ is a direct product of two cyclic groups of order 2, so that $C_2(G)=C^2(G)=1$ and $G=Z_2(G)$. Furthermore G is a direct product of three groups Q, T, U where Q is a quaternion group, T an abelian group all of whose elements not equal to 1 are of order 2, and U an abelian group all of whose elements are of odd order so that in particular G does not contain elements of infinite order.

0.C. If p is a prime number, then a group G is said to be a *p-group* if all its elements are of order a power of p, and *primary components* are [greatest] p-subgroups which contain all the elements of order a power of p.

0.D. If A is an *abelian group*, then its elements of finite order form a subgroup $F(A)$ which is the direct product of its primary components; and $A/F(A)$ does not contain any element not equal to 1 of finite order. The group A may now be said to be of *finite rank*, if

(i) the elements in A which satisfy $x^p=1$, p a prime number, form a finite group,

(ii) there exists a finite number of elements $a(1), \cdots, a(n)$ in A so that

for every element x in A for which $x \not\equiv 1$ modulo $F(A)$ there exist integers $m \neq 0$, $m(i)$ satisfying $\prod_{i=1}^{n} a(i)^{m(i)} \equiv x^m$ modulo $F(A)$.

If $F(A) = 1$, that is, if the identity is the only element of finite order in the abelian group A, then the smallest integer n, meeting the requirement (ii), is termed the *rank* of A. The rank of such a group A is at the same time the greatest number of linearly independent elements in A. If B is a proper subgroup of A, then either the rank of B is smaller than the rank of A; or if A and B have the same rank, then there exists an element w in A, but not in B, and a positive integer i so that w^i is an element in B.

0.E. If G is a *group with abelian central quotient group*, then $C(G) \leqq Z(G)$; and all the properties we are going to use are simple consequences of the following well known formulas:

$$x(yz)x^{-1}(yz)^{-1} = (xyx^{-1}y^{-1})(xzx^{-1}z^{-1}),$$
$$(xy)^i = (yxy^{-1}x^{-1})^{i(i-1)2^{-1}}x^iy^i.$$

0.F. In the introduction we have mentioned a condition, concerning greatest subgroups, which will play a fundamental part in the course of our investigations. For future reference this property may be restated here as follows.

(G) *If S and T are any two subgroups of the group G so that $S < T$ and so that there does not exist any subgroup B between S and T which is different from both S and T, then S is a normal subgroup of T.*

We mention furthermore that the subgroup S of T is said to be a *greatest subgroup* of T, whenever S is a proper subgroup of T and there does not exist any subgroup between S and T which is different from both S and T. Thus (G) implies, in short, that greatest subgroups are normal subgroups.

1. In this section we are going to prove several auxiliary theorems which will be needed in the future. Most of them are concerned with the existence of central elements not equal to 1 which lie in prescribed subgroups, provided the group satisfies the condition (G), enunciated in 0.F.

The crosscut of all the greatest subgroups of a group G is always a characteristic and, therefore, a normal subgroup $M(G)$ of G.

(1.1) *If there exist greatest subgroups of the group G, and if every greatest subgroup of G is a normal subgroup of G, then $G/M(G)$ is an abelian group not equal to* 1.

For if S is a greatest subgroup of G, and if at the same time S is a normal subgroup of G, then G/S is a cyclic group of order a prime number, so that $C(G) \leqq S$.

(1.2) *If the group G is different from 1, if G is generated by a finite number of elements, and if every greatest subgroup of G is a normal subgroup of G, then $G \neq C(G)$.*

Proof. Since G is generated by a finite number of elements, there exists a smallest set N of generators of G (so that no proper subset of N generates G). If u is some element in N, then there exists a greatest subgroup U of G which does not contain u but contains all the other elements in N, since the subgroup of G which is generated by the elements different from u in N is different from G and hence cannot contain u. If V is a subgroup of G such that $U < V$, then V contains u and is therefore equal to G; that is, U is a greatest subgroup of G. Hence $M(G) < G$. But it follows from (1.1) that $C(G) \leqq M(G)$, and this shows that $C(G) < G$.

(1.3) *If G is generated by a finite number of elements, if C is a normal subgroup of G, and if G/C is finite, then C is generated by a finite number of elements.*†

REMARK. That the hypothesis of the finiteness of G/C is really needed may be seen from the fact that the commutator subgroup of a free group with two generators is not generated by a finite number of elements.

Proof. Denote by T some finite set of generators of G and by R a complete set of representatives of G/C so that both R and T are finite sets. If X is any element in G/C, then denote by $r(X)$ the uniquely determined element in R which satisfies $X = Cr(X)$. The elements

$$(X, Y) = r(XY)r(Y)^{-1}r(X)^{-1}$$

for X and Y in G/C and the elements

$$s(t) = tr(Ct)^{-1}$$

for t in T are in the normal subgroup C of G and so are the elements $r(X)s(t)r(X)^{-1}$. Since R and T are finite sets, it follows that the set D consisting of all the elements (X, Y) and $r(X)s(t)r(X)^{-1}$ for X, Y in G/C and t in T is a finite subset of C. Let D^* be the subgroup generated by D.

D^* is clearly a subgroup of C; and G is generated by adjoining the elements $r(X)$ to D^*, since $t = s(t)r(Ct)$. We have

$$\begin{aligned} r(X)(Y, Z)r(X)^{-1} &= r(X)r(YZ)r(Z)^{-1}r(Y)^{-1}r(X)^{-1} \\ &= (X, YZ)r(XYZ)r(Z)^{-1}r(XY)^{-1}(X, Y)^{-1} \\ &= (X, YZ)(XY, Z)^{-1}(X, Y)^{-1}, \end{aligned}$$

† The proof is an obvious application of the Schreier-Reidemeister method.

so that every transform of some (Y, Z) by some $r(X)$ is in D^*. Also

$$r(X)r(Y)s(t)r(Y)^{-1}r(X)^{-1} = (X, Y)r(XY)s(t)r(XY)^{-1}(X, Y)^{-1},$$

so that every transform of any element in D by any of the $r(X)$ is in D^*. Since G is generated by adjoining the elements $r(X)$ to D^*, this shows that D^* is a normal subgroup of G. Since the (X, Y) are in D^*, and since D^* is a subgroup of C, the finite groups G/C and G/D^* are essentially the same; and this implies $D^* = C$. Thus it has been proved that the finite set D of elements is a set of generators of C.

LEMMA 1.4.† *If G is generated by a finite number of elements, if G does not contain elements of infinite order, then every $C_i(G)$ is generated by a finite number of elements; and if, furthermore, condition* (G) *is satisfied by G, then $C_i(G) \neq 1$ implies $C_{i+1}(G) < C_i(G)$.*

Proof. $C_0(G) = G$ is generated by a finite number of elements. Hence assume that it has already been shown that $C_i(G)$ is generated by a finite number of elements. Then $C_i(G)/C_{i+1}(G)$ is generated by a finite number of elements and all its elements are of finite order, since all the elements in $C_i(G) \leqq G$ are of finite order. Hence $C_i(G)/C_{i+1}(G)$ is finite. Since $C_{i+1}(G) = C(C_i(G))$, it follows now from (1.3) that $C_{i+1}(G)$ is generated by a finite number of elements; and hence every $C_i(G)$ is generated by a finite number of elements. It is now a consequence of (1.2) that $C_i(G) \neq 1$ together with condition (G) imply $C_{i+1}(G) < C_i(G)$.

(1.5) *If every greatest subgroup of G is a normal subgroup of G, if the commutator subgroup of G is abelian, if G is generated by a pair of elements a, b so that b is in the commutator subgroup of G, then G is cyclic.*

Proof. G may be generated by adjoining a to $C(G)$. If $c = aba^{-1}b^{-1}$, then the elements $a^i c a^{-i}$ for integral i generate a normal subgroup of G which is contained in $C(G)$—since b permutes with every element in $C(G)$—and modulo which G is abelian, since it contains c, and since G is generated by a and b. Hence $C(G)$ is generated by the elements $a^i c a^{-i}$. This implies however that G is generated by a and c. If c is not a power of a, then there exists a greatest subgroup of G which contains a, but not c. This subgroup V is a greatest subgroup of G so that $C(G) \leqq V$. Since V contains both a and $C(G)$, it follows that $V = G$ so that V would contain c in contradiction to the choice of V. Hence c is a power of a so that G is the cyclic group, generated by a.

(1.6) *Suppose that the commutator subgroup of the group G is abelian and that G is generated by two elements a, b where $bc = cb$ for every c in $C(G)$. Let*

† Cf. Corollary 3.7 below which improves this result in certain respects.

$b_0=b$, $b_{i+1}=ab_ia^{-1}b_i^{-1}$; and let $G_0=G$ and G_{i+1} be generated by adjoining a to $C(G_i)$. Then G_i is generated by a and b_i, $C(G_i)$ by the elements $a^jb_{i+1}a^{-j}$ for integral j, and $C(G_i)=C^{i+1}(G)$.

Proof. Since $G_0=G$, $b_0=b$, G_0 is generated by a and b_0. Assume now that G_i is generated by a and b_i. Let V_i be the subgroup generated by the elements $a^jb_{i+1}a^{-j}$ for integral j. Then $V_i\leqq C(G_i)$. Since b permutes with every element in $C(G)$ and therefore with every element in V_i, since the set of generators of V_i is transformed into itself by a, and since G is generated by a and b, it follows that V_i is a normal subgroup of G and therefore of G_i. Since finally G_i/V_i is abelian, as b_{i+1} is in V_i, it follows that $C(G_i)\leqq V_i$, that is, $C(G_i)=V_i$. Since G_{i+1} is generated by adjoining a to $C(G_i)=V_i$, it follows that G_{i+1} is generated by a and b_{i+1}.

Since $C(G_0)=C(G)=C^1(G)$, we may assume that $V_i=C(G_i)=C^{i+1}(G)$. Since b permutes with every element in $C^{i+1}(G)$, it follows that $C^{i+2}(G)$ is generated by the elements $a^jxa^{-j}x^{-1}$ for integral j and x in $C^{i+1}(G)$. Since $C^{i+1}(G)$ is an abelian group which is generated by the elements $a^rb_{i+1}a^{-r}$ for integral r, it follows that $C^{i+2}(G)$ is generated by the elements $a^ra^sb_{i+1}a^{-s}b_{i+1}^{-1}a^{-r}$ for integral r and s. If $s<0$, then

$$\begin{aligned}(a^sb_{i+1}a^{-s}b_{i+1}^{-1})^{-1} &= b_{i+1}a^sb_{i+1}^{-1}a^{-s}\\ &= a^s(a^{-s}b_{i+1}a^sb_{i+1}^{-1})a^{-s};\end{aligned}$$

and if $0\leqq s$, then

$$\begin{aligned}a^{s+1}b_{i+1}a^{-1-s}b_{i+1}^{-1} &= a^sab_{i+1}a^{-1}b_{i+1}^{-1}a^{-s}a^sb_{i+1}a^{-s}b_{i+1}^{-1}\\ &= a^sb_{i+2}a^{-s}a^sb_{i+1}a^{-s}b_{i+1}^{-1}\end{aligned}$$

and it follows by complete induction with regard to s that we have

$$a^sb_{i+1}a^{-s}b_{i+1}^{-1}=\prod_{j=0}^{s-1}a^jb_{i+2}a^{-j} \qquad \text{for } 0<s.$$

Hence $C^{i+2}(G)\leqq V_{i+1}=C(G_{i+1})\leqq C^{i+2}(G)$, and this completes the proof.

LEMMA 1.7. *If the commutator subgroup $C(G)$ of the group G is abelian, if G is generated by two elements u and v so that $vc=cv$ for every c in $C(G)$, and if G satisfies* (G), *then either $C^i(G)=1$ or $C^{i+1}(G)<C^i(G)$.*

Proof. As in (1.6) we use the following notations: $v_0=v$, $v_{i+1}=uv_iu^{-1}v_i^{-1}$; $G_0=G$, G_{i+1} is the subgroup generated by adjoining u to $C(G_i)$. Then it follows from (1.6) that *G_i is generated by u and v_i, and that $C(G_i)=C^{i+1}(G)$.*

Assume now that for some integer i we have $C^i(G)=C^{i+1}(G)$. Then $C(G_{i-1})=C(G_i)$ and therefore $G_i=G_{i+1}$. Hence G_i is generated by u and v_{i+1},

where v_{i+1} is contained in the abelian commutator subgroup $C(G_i) \leqq C(G)$ of G_i. Since G satisfies (G), so does G_i. Hence it follows from (1.5) that G_i is a cyclic group. Consequently $C(G_i)=1$, $v_{i+1}=1$, so that G_i is the cyclic group generated by u. Thus it follows that $1=C(G_i)=C^{i+1}(G)=C^i(G)$, and this completes the proof.

(1.8) *If B is an abelian group all of whose elements not equal to* 1 *are of the same order, if B possesses a countable basis* $\cdots, b(i), b(i+1), \cdots$, *and if the group G is generated by adjoining an element g to B that is subject to the relation $g^{-1}b(i)g=b(i+1)$, then condition* (G) *is not satisfied by the group G.*

Proof. B is a normal subgroup of G, and G/B is an infinite cyclic group. The elements not equal to 1 in B are either all of order a prime number p or they are all of infinite order. Accordingly three cases are distinguished.

CASE 1. The elements not equal to 1 in B are of order 2.

Denote by U the subgroup of B which is generated by all the elements $b(i)b(i+1)b(i+2)$, and by S the subgroup of G which is generated by $b(1)b(2)b(3)$ and g. Then U is the crosscut of B and S.

If $x \neq 1$ is any element in U, then it begins with some $b(i)$ and ends with some $b(i+j)$ where $1<j$. Thus U does not contain $b(1)b(2)$. Since $b(1)gb(1)^{-1}=b(1)b(2)g$, this shows that S is not a normal subgroup of G.

Suppose now that $S<T \leqq G$. Then T contains elements w which are in B but not in U. Such an element has a certain "length," so that T contains an element w in B but not in U which is of shortest length. Suppose that w begins with $b(i)$ and ends with $b(i+j)$. Since U contains $b(i+j-2)b(i+j-1)b(i+j)$, it follows that $0 \leqq j \leqq 1$. For $j=1$ we would find however that $b(i)b(i+1)b(i)b(i+1)b(i+2)=b(i+2)$ would be of shorter length. Hence $w=b(i)$ and $T=G$, so that S is a greatest subgroup of G though not a normal one.

CASE 2. The elements not equal to 1 in B are of order a prime number $p \neq 2$.

Denote by U the subgroup of B which is generated by the elements $b(i)b(i+1)$ and by S the subgroup of G which is generated by g and $b(1)b(2)$. Then U is the crosscut of S and B.

Any element $x \neq 1$ in U contains at least two elements $b(i)$ and $b(j) \neq b(i)$ as factors. Thus $b(1)$ is no element in U. Since 2 is relatively prime to p, it follows that $b(1)^2=b(1)b(2)^{-1}b(1)b(2)$is not an element in U. Hence $b(1)b(2)^{-1}$ is not contained in U. Since $b(1)gb(1)^{-1}=b(1)b(2)^{-1}g$, this shows that S is not a normal subgroup of G.

Suppose now that $S<T \leqq G$. Then T contains elements w which are contained in B but not in U. Amongst these there are some of shortest "length"; and since U contains $b(i)b(i+1)$, it follows that such an element of shortest

length has the form $b(i)^n$ where n is relatively prime to p. Hence T contains $b(i)$ and therefore every $b(j)$, so that $T=G$. Thus S is a greatest subgroup of G but not a normal one.

CASE 3. The elements not equal to 1 in B are of infinite order.

Then let p be some prime number. Clearly the subgroup B^p which is generated by the elements $b(i)^p$ is a normal subgroup of G. The group G/B^p is then just of the type discussed under Cases 1 and 2. Hence G/B^p does not satisfy (G), so that G itself cannot satisfy (G).

LEMMA 1.9. *If condition* (G) *is satisfied by the group G, if $H\neq 1$ is a normal subgroup of G, if H is abelian and G/H cyclic, then the crosscut of $Z(G)$ and H is different from* 1.

Proof. Let g be an element that generates G modulo H. Then the crosscut of $Z(G)$ and H consists of exactly those elements x in H which satisfy $xg=gx$.

CASE 1. H contains elements not equal to 1 of finite order.

Then H contains an element u of order a prime number p. Denote by U the subgroup generated by u and g. The crosscut V of U and H is generated by the elements $u(i)=g^{-i}ug^i$ for integral i.

If there did not exist any relation between the elements $u(i)$, then U would be a group of the type discussed in (1.8). This is impossible, since G and U satisfy condition (G). Hence there exists a relation between the $u(i)$. This proves that V is generated by a finite number of elements, so that V is a finite abelian group not equal to 1.

There is nothing to prove if $V\leqq Z(U)$. If this is not the case, then $1<C(U)\leqq V$. Since V is finite, it follows that the chain of subgroups $C^i(U)$ ends, so that there exists an integer i satisfying $C^{i+1}(U)=C^i(U)<C^{i-1}(U)\leqq V$. Since G and U satisfy (G), it follows from Lemma 1.7 that $C^i(U)=1$. Hence $1<C^{i-1}(U)\leqq Z(U)$ so that the crosscut of $Z(U)$ and V is different from 1. Thus the crosscut of $Z(G)$ and H is different from 1.

CASE 2. All the elements not equal to 1 in H are of infinite order.

Then there exists for every element x in H and for every integer n at most one element y in H so that $x=y^n$.

Let u be an element not equal to 1 in H. Denote by U the subgroup generated by u and g. The crosscut V of U and H is generated by the elements $b(i)=g^{-i}ug^i$ for integral i.

If there did not exist any relation between the elements $b(i)$, then U would be of the type discussed in (1.8). This is impossible, since G and U satisfy condition (G). Hence there exists some relation between the $b(i)$, and we may assume without loss in generality that this relation has the following form:

$$(\alpha) \qquad \prod_{i=1}^{n-1} b(i)^{m(i)} = b(n)^m \qquad \text{with } m(1)m \not\equiv 0;$$

though there does not exist any relation between $b(1), \cdots, b(n-1)$. Denote now by p some prime number that does not divide m or $m(1)$ and denote by V^p the subgroup of V which is generated by the elements x^p for x in V. Clearly V^p is a normal subgroup of U.

One verifies that $b(n+i)^{m^{i+1}}$ for $0<i$ is contained in the subgroup generated by $b(1), \cdots, b(n-1)$. Since

$$b(0)^{m(1)} = \prod_{i=2}^{n-1} b(i-1)^{-m(i)} b(n-1)^m,$$

as follows on transforming (α) by g^{-1}, one verifies too that $b(0)^{m(1)}$, $b(-i)^{m(1)i+1}$ (for $0<i$) are elements in the subgroup generated by $b(1), \cdots, b(n-1)$.

Suppose now that $r=\prod_{i=1}^{n-1} b(i)^{r(i)} \equiv 1$ modulo V^p. Hence there exists an element s in V so that $r=s^p$. Then s may be represented as a product of elements $b(j)$ with $-i \leqq j \leqq i$ for $0<i$. Thus $s^{m(1)i+1 m^{i+1}}$ is certainly an element in the subgroup generated by $b(1), \cdots, b(n-1)$. If we put $t=m(1)^{i+1}m^{i+1}$, then t is relatively prime to p; and $r^t=s^{tp}$. Suppose that $s^t=\prod_{i=1}^{n-1} b(i)^{s(i)}$. Then

$$\prod_{i=1}^{n-1} b(i)^{tr(i)} = \prod_{i=1}^{n-1} b(i)^{ps(i)}.$$

Since there does not exist any relation between $b(1), \cdots, b(n-1)$, it follows that $tr(i)=ps(i)$. Since t is relatively prime to p, this implies that $r(i) \equiv 0$ modulo p. Thus we have proved that $b(1), \cdots, b(n-1)$ *form a basis of* V *modulo* V^p. This shows in particular that $V/V^p \neq 1$ is a finite abelian group. Hence it follows from what has been proved under Case 1 that the crosscut of $Z(U/V^p)$ and V/V^p is different from 1, since G and therefore U/V^p satisfy (G), since U/V is cyclic, and since V/V^p is abelian.

The automorphism which is induced by g in V/V^p has therefore fixed-elements not equal to 1; and it may be represented by the matrix

$$\begin{pmatrix} 0 & 1 & 0 & 0 & \cdots & 0 & 0 \\ 0 & 0 & 1 & 0 & \cdots & 0 & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdots & \cdot & \cdot \\ 0 & 0 & 0 & \cdot & \cdots & 0 & 1 \\ m(1)m^{-1} & m(2)m^{-1} & m(3)m^{-1} & \cdot & \cdots & m(n-2)m^{-1} & m(n-1)m^{-1} \end{pmatrix}$$

so that the determinant

$$\begin{vmatrix} -1 & 1 & 0 & 0 & \cdots & 0 & 0 \\ 0 & -1 & 1 & 0 & \cdots & 0 & 0 \\ \cdot & \cdot & \cdot & \cdot & \cdots & \cdot & \cdot \\ 0 & 0 & 0 & 0 & \cdots & -1 & 1 \\ m(1)m^{-1} & m(2)m^{-1} & m(3)m^{-1} & \cdot & \cdots & m(n-2)m^{-1} & m(n-1)m^{-1}-1 \end{vmatrix} = D,$$

whose coefficients are rational numbers with denominators prime to p, is divisible by p; that is, the integer Dm^{n-1} is divisible by p. Since this holds true for every prime number p that does not divide m or $m(1)$, and since there exists an infinity of such prime numbers, it follows that $D=0$. Consequently there exists an element different from 1 in V which is a fixed-element under the automorphism induced by g in V, so that the crosscut of $Z(U)$ and V is different from 1. It follows that the crosscut of $Z(G)$ and H is different from 1.

LEMMA 1.10. *If the normal subgroup H of the group G is different from* 1, *if H is abelian and of finite rank, if there exists for every subgroup $K \neq 1$ of H and for every element g in G such that $gK = Kg$ an element $h \neq 1$ in K such that $gh = hg$, and if there exists an ascending chain* of subgroups B_ν such that*

(1) *$H = B_2$;*

(2) *B_ν is a normal subgroup of $B_{\nu+1}$;*

(3) *$B_{\nu+1}/B_\nu$ is abelian;*

(4) *B_ν is for limit-ordinals ν the set of all the elements contained in subgroups B_μ for $\mu < \nu$;*

(5) *$G = B_\tau$,*

then the crosscut of H and $Z(G)$ is different from 1.

Proof. It is possible to well-order the elements in G in such a way: $g(1), g(2), \cdots, g(\kappa), g(\kappa+1), \cdots$ that all the elements in B_ν precede all the elements that are not contained in B_ν.

If H contains elements of finite order that are different from 1, then there exists a prime number p such that H contains elements of the exact order p. The subgroup generated by these elements is finite, since H is of finite rank; and this subgroup shall be denoted by B_1. If however H does not contain elements of finite order which are different from 1, then we put $H = B_1$ so that B_1 is an abelian group different from 1 which is either finite or does not contain any element of finite order except 1 and which is a normal subgroup of G, since it is a characteristic subgroup of the normal subgroup H of G.

Denote by $H(\nu)$ the set of elements x in B_1 which satisfy $g(\nu)x = xg(\nu)$.

* If we use the terminology introduced in §3, this amounts to saying that G/H is metacyclic.

Then it follows from our hypothesis concerning H and G that every $H(\nu)$ is a subgroup, different from 1, of B_1.

If $1<\mu$, then denote by $K(\mu)$ the crosscut of all the $H(\nu)$ for $\nu<\mu$. The $K(\mu)$ are subgroups of B_1 too.

If B_1 is finite, then the $K(\mu)$ form a descending chain of finite abelian groups. If B_1 is infinite, then 1 is the only element of finite order in B_1 so that there exists to every integer i and to every b in B_1 at most one solution x of $x^i=b$ in B_1. If g is an element in G, b an element in B_1, i an integer, then it follows that $gb^i=b^ig$ implies $b^i=gb^ig^{-1}=(gbg^{-1})^i$ and hence $gb=bg$ too. The $K(\mu)$ form therefore a descending chain of subgroups of the group B_1 of finite rank with the property that $K(\rho)<K(\sigma)$ if, and only if, the rank of $K(\rho)$ is smaller than the rank of $K(\sigma)$.

Thus it follows that there exists in both cases a smallest ordinal τ so that $K(\tau)=K(\nu)$ for $\tau<\nu$, that is, $K(\tau)\leqq Z(G)$. Since $K(\tau)<K(\mu)$ for $\mu<\tau$, it follows that τ cannot be a limit-ordinal so that $\tau=\rho+1$, and so that $K(\rho)\neq 1$. There exists furthermore a uniquely determined ordinal δ so that $g(\rho)$ is an element in $B_{\delta+1}$, but not in B_δ. Then all the $g(\nu)$ for $\nu<\rho$ are contained in $B_{\delta+1}$, so that all the commutators $g(\nu)^{-1}g(\rho)^{-1}g(\nu)g(\rho)$ are elements in B_δ; that is, they are elements $g(\mu)$ for $\mu<\rho$. If x is an element in $K(\rho)$, and if $\nu<\rho$, then

$$\begin{aligned} g(\nu)g(\rho)xg(\rho)^{-1} &= g(\rho)g(\nu)g(\nu)^{-1}g(\rho)^{-1}g(\nu)g(\rho)xg(\rho)^{-1} \\ &= g(\rho)g(\nu)xg(\nu)^{-1}g(\rho)^{-1}g(\nu)g(\rho)g(\rho)^{-1} \\ &= g(\rho)xg(\nu)g(\nu)^{-1}g(\rho)^{-1}g(\nu)g(\rho)g(\rho)^{-1} \\ &= g(\rho)xg(\rho)^{-1}g(\nu) \end{aligned}$$

so that $g(\rho)K(\rho)=K(\rho)g(\rho)$. Hence it follows from $K(\rho)\neq 1$ that there exists an element $v\neq 1$ in $K(\rho)$ satisfying $g(\rho)v=vg(\rho)$ and this proves that $K(\tau)\neq 1$. Since $K(\tau)\leqq Z(G)$ and since $K(\tau)\leqq B_1\leqq H$, it follows that the crosscut of H and $Z(G)$ is different from 1.

LEMMA 1.11. *Suppose that the property* (P) *of groups satisfies the following conditions*:

(i) *If* (P) *is satisfied by the group* H, *then* (P) *is satisfied by every quotient group of* H;

(ii) *If* (P) *is satisfied by the group* $H\neq 1$, *then* $Z(H)\neq 1$.

Then there exists for every group G, *satisfying* (P), *an ordinal* ζ *such that* $G=Z_\zeta(G)$.

Proof. If (P) is satisfied by the group G, then (P) is satisfied by $G/Z_\nu(G)$. If $Z_\nu(G)<G$, then it follows from (ii) that the central of $G/Z_\nu(G)$ is different

from 1, that is, $Z_\nu(G) < Z_{\nu+1}(G)$. Since the $Z_\nu(G)$ form an ascending chain of subgroups, there exists an ordinal ζ such that $Z_\zeta(G) = Z_{\zeta+1}(G)$ and consequently $G = Z_\zeta(G)$.

2. In this section we shall give a characterization of the groups G which satisfy $G = Z_\zeta(G)$ for some ordinal ζ.

LEMMA 2.1. *If $G = Z_\zeta(G)$, and if S is a subgroup of G, then $S = Z_\zeta(S)$.*

For one verifies by complete (transfinite) induction that the crosscut of S and $Z_\nu(G)$ is part of $Z_\nu(S)$.

It is a consequence of Jordan-Hölder's theorem that a finite group G is soluble if, and only if, the series of subgroups $C_i(G)$ ends with 1. Thus the following definition of solubility coincides for finite groups with the classical concept.

The (finite or infinite) group G is termed *soluble* whenever there exists an ascending chain of subgroups B_ν for ordinals ν with $0 \leqq \nu \leqq \beta$ so that

(a) $B_0 = 1$;

(b) *B_ν is a normal subgroup of G;*

(c) *$B_{\nu+1}/B_\nu$ is an abelian group of finite rank;*

(d) *B_ν is for limit-ordinals ν the set of all the elements contained in groups B_μ for $\mu < \nu$;*

(e) $B_\beta = G$.

If $G \neq 1$, then there exists a first $B_\nu \neq 1$, so that we may assume without loss in generality that $B_1 \neq 1$.

THEOREM 2.2. *There exists an ordinal ζ so that $G = Z_\zeta(G)$ if, and only if, G is a soluble group, satisfying condition* (G).

Proof. We assume first the existence of an ordinal ζ so that $G = Z_\zeta(G)$. Since every $Z_{\nu+1}(G)/Z_\nu(G)$ is abelian, there exists an ascending chain of subgroups $B_{\nu,\sigma}$ so that $Z_\nu(G) = B_{\nu,0} \leqq B_{\nu,\rho} \leqq B_{\nu,\sigma} \leqq B_{\nu,\tau} = Z_{\nu+1}(G)$ for $\rho < \sigma$, and $B_{\nu,\rho+1}/B_{\nu,\rho}$ is a cyclic group, and so that condition (d) is satisfied by the $B_{\nu,\sigma}$. Since every subgroup of the central of a group is a normal subgroup, it follows that all the $B_{\nu,\sigma}$ are normal subgroups of G; and this proves the solubility of G.

Suppose now that S is a subgroup of G and that T is a greatest subgroup of S. Since $Z_0(S) = 1$, and since T contains $Z_\nu(S)$ whenever ν is a limit-ordinal and T contains all the $Z_\mu(S)$ for $\mu < \nu$, there exists a greatest subscript τ so that $Z_\tau(S) \leqq T$; and since $T < S$ and $S = Z_\zeta(S)$ by Lemma 2.1, it follows that $Z_\tau(S) < S$. Hence there exists an element w in $Z_{\tau+1}(S)$ that is not contained in T. If s is any element in S, then $wsw^{-1}s^{-1}$ is an element in $Z_\tau(S)$ and therefore in T, so that $wT = Tw$. Since T is a greatest subgroup of S, S is generated by

adjoining w to T; and this shows that T is a normal subgroup of G, so that (G) is satisfied by the group G.

Assume now conversely that G is a soluble group and (G) is satisfied by G. Let the B_ν form an ascending chain of subgroups of G which satisfies conditions (a) to (e) and in addition $B_1 \neq 1$. B_1 is a normal abelian subgroup of G. Suppose that K is a subgroup not equal to 1 of B_1 and g an element in G so that $gK = Kg$. Then it follows from (G) and Lemma 1.9 that there exist elements $v \neq 1$ in K so that $gv = vg$. Hence it follows from Lemma 1.10 that $Z(G) \neq 1$.

If G is a soluble group satisfying (G), then every quotient group of G is a soluble group satisfying (G), as may be seen by applying, to the ascending chain B_ν, the homomorphism which defines the quotient group. Since soluble groups not equal to 1, satisfying (G), have a central different from 1, it follows from Lemma 1.11 that there exists an ordinal ζ so that $G = Z_\zeta(G)$.

COROLLARY 2.3. *If G is a finite group, then each of the following properties implies the others:*

(1) *There exists an integer h so that $G = Z_h(G)$.*

(2) *G is a direct product of p-groups.*

(3) *Condition* (G) *is satisfied by the group G.*

The equivalence of (1) and (2) is a well known theorem.* That (3) is a consequence of (1), follows from Theorem 2.2. If finally (3) is satisfied, then it follows from the finiteness of G and from condition (G) that G is soluble by Lemma 1.4; and it follows from Theorem 2.2 that (1) is satisfied by G.†

COROLLARY 2.4. *If* (G) *is satisfied by the group G, and if there exists a finite ascending chain of subgroups B_i so that*

(i) $B_0 = 1$,

(ii) *B_{i-1} is a normal subgroup of B_i and B_i/B_{i-1} is an abelian group of finite rank,*

(iii) $B_k = G$,

then $G = Z_\zeta(G)$ for some (finite or transfinite) ordinal ζ, and each quotient group $Z_{\nu+1}(G)/Z_\nu(G)$ is an abelian group of finite rank.

* Cf. for example, Burnside, *Theory of Groups of Finite Order*.

† It is known (cf. for example H. Wielandt, Mathematische Zeitschrift, vol. 41 (1936), pp. 281–282, or H. Zassenhaus, *Lehrbuch der Gruppentheorie*, vol. 1, p. 108) that for finite groups condition (G) is a consequence of the weaker condition that every greatest subgroup of the whole group be a normal subgroup. Such a condition would be fairly ineffective in the case of infinite groups, since infinite groups need not possess any greatest proper subgroups.

A fact that amounts essentially to the equivalence of (1) and (3) has been proved recently by H. Wielandt, Mathematische Zeitschrift, vol. 45 (1939), pp. 209–244, (16). His condition that every subgroup be "subnormal" (nachinvariant) implies (G) and is equivalent to (G) only in case a suitable chain condition is satisfied; cf. also Theorem 4.13 and Corollary 4.15 below.

Proof. We show by complete induction with regard to i that each of the subgroups B_i fulfills the assertions of our statement. This is clearly true for B_1, since B_1 is an abelian group of finite rank. Thus we may assume that it holds true for B_{i-1}. Since B_{i-1} is a normal subgroup of B_i, and since the $Z_\nu(B_{i-1})$ are characteristic subgroups of B_{i-1}, they are normal subgroups of B_i. Since $Z_{\nu+1}(B_{i-1})/Z_\nu(B_{i-1})$ and $B_i/B_{i-1}=B_i/Z_\rho(B_{i-1})$ are abelian groups of finite rank—by the induction hypothesis—this shows that B_i is a soluble group which satisfies (G); and it follows from Theorem 2.2 that $B_i=Z_\sigma(B_i)$ for some ordinal σ. The subgroup of those classes of $Z_{\nu+1}(B_i)/Z_\nu(B_i)$ which are represented by elements in B_{i-1} is isomorphic with a subgroup of $Z_{\nu+1}(B_{i-1})/Z_\nu(B_{i-1})$ and is therefore an abelian group of finite rank; and $Z_{\nu+1}(B_i)/Z_\nu(B_i)$ is modulo this subgroup isomorphic to a subgroup of B_i/B_{i-1} —an abelian group of finite rank. Thus $Z_{\nu+1}(B_i)/Z_\nu(B_i)$ is an abelian group of finite rank; and this completes the proof.

That it is impossible to prove $G=Z_h(G)$ for integral h may be seen from the following:

EXAMPLE 2.5. *Let B be an abelian group of type p^∞, that is, B is generated by a sequence of elements $b(i)$ so that $b(1)$ is an element of order p, $b(i)^p=b(i-1)$. An automorphism γ of B is defined by $x^\gamma=x^{1+p}$.*

Let G be the group which is generated by adjoining to B an element u, subject to the relation $u^{-1}xu=x^\gamma$ for x in B.

B is the commutator subgroup of G, and G/B is an infinite cyclic group. If i is a positive integer, then $Z_i(G)$ is generated by $b(i)$, so that $B=Z_\omega(G)$, $G=Z_{\omega+1}(G)$. Thus it follows from Theorem 2.2 that (G) is satisfied by G. Finally it may be of interest* to note that $B=C(G)=C^2(G)$.

LEMMA 2.6. *Suppose the subgroups W_ν for $0\leqq\nu\leqq\tau$ of the group G satisfy*:

(a) $W_0=1$;

(b) $W_\nu\leqq W_\mu$ *for* $\nu<\mu$;

(c) W_ν *is for limit-ordinals ν the set of all the elements contained in W_μ for $\mu<\nu$*;

(d) W_ν *is a normal subgroup of G*;

(e) *if the subgroup S of G contains W_ν, and if w is an element in $W_{\nu+1}$, then $wS=Sw$.*

---

* For one proves easily, using an argument of Philip Hall, that if $G=Z_h(G)$ for some integer h, then $C^h(G)=1$; and if $C^k(G)=1$ for some integer k, then $G=Z_k(G)$. That neither part of this statement holds true when infinite indices are admitted, may be seen from the following example that has been discussed in the proof of (1.8). Let G be the group generated by elements $b(i)$ for integral i (positive or negative or zero) and by an element g, subject to the relations $b(i)b(j)=b(j)b(i)$, $b(i)^2=1$, $g^{-1}b(i)g=b(i+1)$. $C(G)$ is abelian, $Z(G)=1$, (G) is not satisfied by G, but the crosscut of the groups $C^i(G)$ for positive integers i consists of the identity only.

Then (G) *is satisfied by the group G, if (and only if)* (G) *is satisfied by the quotient group* G/W_τ.

Proof. Suppose that S is a subgroup of G, and that T is a greatest subgroup of S. Denote by S_ν the crosscut of S and W_ν, and by T_ν the crosscut of T and W_ν. If ν is a limit-ordinal, and if $S_\mu = T_\mu$ for every $\mu < \nu$, then $S_\nu = T_\nu$ by (c). Since $S_0 = T_0 = 1$, there exists therefore a greatest ordinal ρ so that $S_\rho = T_\rho$. If $\rho = \tau$, then T/T_τ is a greatest subgroup of S/S_τ, and S/S_τ is essentially a subgroup of G/W_τ—a group which satisfies (G)—so that T is a normal subgroup of S. If $\rho < \tau$, then there exists an element w in $S_{\rho+1}$ which is not contained in T. It is a consequence of (e) that there exists for every element g in G an integer i and an element v in W_ρ so that $w^{-1}gw = g^i v$. If g is in particular an element in S, then v is an element in S, and therefore an element in $S_\rho = T_\rho \leqq T$. Thus $wT = Tw$. Since w is not in T, S is generated by adjoining w to T, and hence T is a normal subgroup of S.

The *norm* $N(G)$ of the group G has been defined* as the set of all the elements g in G which satisfy $gS = Sg$ for every subgroup S of G. The iterated norms $N_\nu(G)$ may be defined as usual. An obvious consequence of Lemma 2.6 is the fact that (G) is satisfied by a group G, whenever $G = N_\kappa(G)$ for some ordinal κ. Using Corollary 2.3 this implies the following statement:

If G is a finite group, then properties (1) *to* (3) *of Corollary* 2.3 *are equivalent to the following condition*:

(4) $G = N_r(G)$ *for some integer* r.

If S is a greatest abelian subgroup of the group G, then it is clear that S equals its centralizer in G. It is furthermore known† that every greatest normal and abelian subgroup of G equals its centralizer in G, if only $G = Z_h(G)$ for some integer h. This may be generalized as follows.

If S is a greatest normal and abelian subgroup of the group G, and if $G = Z_\zeta(G)$ *for some (finite or infinite) ordinal* ζ, *then the centralizer of S in G is equal to S.*

Proof. Denote by T the centralizer of S in the group G. Since S is normal and abelian, $S \leqq T$ and T is a normal subgroup of G. Denote by S_ν the crosscut of S and $Z_\nu(G)$; and denote by T_ν the crosscut of T and $Z_\nu(G)$. Since $S_0 = T_0 = Z_0(G) = 1$, we may assume that $S_\mu = T_\mu$ has been proved for every $\mu < \nu$.

CASE 1. $\nu = \rho + 1$.

Then $S_\rho = T_\rho$. If u is any element in T_ν, s an element in G, then $usu^{-1}s^{-1}$

* Cf. Compositio Mathematica, vol. 1 (1934), pp. 254–283.

† Cf. H. Zassenhaus, *Lehrbuch der Gruppentheorie*, vol. 1, p. 108.

is an element in $Z_\rho(G)$ as well as in T; and therefore it is an element in T_ρ. Hence it is an element in S. Thus the subgroup U, generated in adjoining u to S, is a normal subgroup of G. Since u is an element in T, U is abelian. Hence $U=S$ so that $T_\nu \leqq S$ or $S_\nu = T_\nu$.

CASE 2. ν is a limit-ordinal.

Then S_ν is the join of the S_μ for $\mu<\nu$; and T_ν is the join of the T_μ for $\mu<\nu$; and this implies that $S_\nu = T_\nu$.

Thus it follows in particular that $S=S_\xi=T_\xi=T$, since $G=Z_\xi(G)$.

3. In this section we are going to apply the results of the preceding sections on groups without elements of infinite order; in particular we shall be interested in the possibility of decomposing a group into a direct product of p-groups.

3.A. For some of our purposes it will be sufficient to assume instead of solubility a somewhat weaker property. A group G may be termed *metacyclic*, if there exists an ascending chain of subgroups B_ν with the properties:

(a) $B_0=1$;

(b) *B_ν is a normal subgroup of $B_{\nu+1}$ and $B_{\nu+1}/B_\nu$ is abelian*;

(c) *B_ν is for limit-ordinals ν the set of all the elements which are contained in groups B_μ for $\mu<\nu$*;

(d) $B_\beta=G$.

If $G\neq 1$, then there exists a first $B_\nu\neq 1$, so that we may assume without loss in generality that $B_1\neq 1$. One verifies readily that every abelian group possesses an ascending chain of subgroups B_ν satisfying (a), (c), (d) and instead of (b) the stronger condition that $B_{\nu+1}/B_\nu$ be cyclic. This shows that there is no loss of generality in assuming that the chain B_ν which defines G as a metacyclic group satisfies (a), (c), (d) and

(b'') *B_ν is a normal subgroup of $B_{\nu+1}$ and $B_{\nu+1}/B_\nu$ is cyclic.*

If there exists a chain of subgroups B_ν satisfying (a), (b''), (c) and (d), then G is said to be *metacyclic of length less than or equal to β.*

Soluble groups are metacyclic; and these two concepts coincide for finite groups. But it is easy to construct examples of infinite metacyclic p-groups which are not soluble, though they satisfy condition (G); cf. Example 3.4 below.

LEMMA 3.1. *Every metacyclic group without elements of infinite order which is generated by a finite number of elements is finite.*

Proof. If G is metacyclic, then there exists an ascending chain of subgroups B_ν which satisfies the above conditions (a) to (d). We may assume without loss of generality that $B_\mu<B_\nu$ for $\mu<\nu$. If the length of the chain

is 1, then the group is abelian; and abelian groups without elements of infinite order are finite, if they are generated by a finite number of elements. Hence we may assume that our statement holds true for metacyclic groups possessing a chain of shorter length. Since G is generated by a finite number of elements, it is impossible that the length β be a limit-ordinal, since otherwise there would exist an ordinal $\gamma<\beta$ so that all the elements in the finite set of generators of G are contained in B_γ and so that therefore $B_\beta=G=B_\gamma$ for $\gamma<\beta$ which contradicts our hypothesis concerning the chain B_ν. Hence $\beta=\alpha+1$ and B_α is a normal subgroup of G. Since G/B_α is an abelian group without elements of infinite order which is generated by a finite number of elements, it is a finite group; and it follows from (1.3) that B_α is generated by a finite number of elements. It is therefore a consequence of the induction hypothesis that B_α is finite; and G is finite, since its quotient group, modulo a finite normal subgroup, is finite.

THEOREM 3.2. *Each of the following three properties of a group G without elements of infinite order implies the others.*

(1) $G=Z_\zeta(G)$ *for some ordinal* ζ.

(2) G *is a direct product of soluble p-groups.*

(3) G *is soluble and satisfies condition* (G).

REMARK. It may be verified easily that all the elements of a soluble group are of finite order if, and only if, there exists an ascending chain of subgroups B_ν satisfying conditions (a) to (e) of §2 and in addition:

(c′) $B_{\nu+1}/B_\nu$ *is a finite abelian group.*

Proof. The equivalence of (1) and (3) has been established as Theorem 2.2. In order to prove the equivalence of (1) and (2) we show first

(3.2.1) *If G satisfies* (1), *and if the subgroup S of G is generated by a finite number of elements, then S is a finite group and a direct product of p-groups.*

It is a consequence of Lemma 2.1 that there exists an ordinal σ so that $S=Z_\sigma(S)$; and it is a consequence of Theorem 2.2 that S is soluble and metacyclic. Hence it follows from Lemma 3.1 that S is finite; and it is a consequence of Corollary 2.3 that S is a direct product of p-groups.

Denote now by u and v two elements of order a power of the same prime number p. The subgroup S, generated by u and v, is by (3.2.1) the direct product of its primary components and is consequently a p-group. Hence uv is an element of order a power of p. This shows in particular that the set G_p of all the elements in G whose order is a power of p is a subgroup of G; and one verifies now as usual that G is the direct product of its primary components G_p.

Since G_p is a subgroup of G, there exists by Lemma 2.1 some ordinal α so that $G_p = Z_\alpha(G_p)$. This implies by Theorem 2.2 that G_p is soluble. Thus (2) is a consequence of (1).

In order to prove that (1) is a consequence of (2) we show first

(3.2.2) *If G is a soluble p-group, then there exists an ordinal ζ so that $G = Z_\zeta(G)$.*

Since $G \neq 1$ is a soluble p-group, there exists an ascending chain of subgroups B_ν which satisfies conditions (a) to (e) of §2 and in addition the condition that B_1 is a finite abelian p-group not equal to 1.

Suppose now that $K \neq 1$ is a subgroup of B_1, and that g is an element in G so that $gK = Kg$. Then g induces in K an automorphism whose order is a power of p; and one proves as usual* that such an automorphism of K possesses fixed-elements not equal to 1; that is, there exist elements $v \neq 1$ in K so that $gv = vg$. Hence the central of a soluble p-group is different from 1 whenever the group is different from 1, as follows from Lemma 1.10.

Quotient groups of soluble p-groups are soluble p-groups. Hence (3.2.2) is a consequence of Lemma 1.11 and the fact just proved.

Suppose now that (2) is satisfied by (G). Then G is the direct product of soluble p-groups H_ν. To every H_ν there exists by (3.2.2) an ordinal $\zeta(\nu)$ so that $H_\nu = Z_{\zeta(\nu)}(H_\nu)$. Let ζ be some ordinal which is greater than all the $\zeta(\nu)$. Since it may be proved by complete (transfinite) induction that $Z_\rho(G)$ is the direct product of the groups $Z_\rho(H_\nu)$, it follows in particular that $G = Z_\zeta(G)$ so that (1) is a consequence of (2).

The theorem that nilpotent groups without elements of infinite order are direct products of their primary components has been proved by deriving (3.2.1) and by applying Corollary 2.3, that is, by reduction to the finite case. An *alternative proof* may be offered which does not make use of this reduction and which consequently includes a proof of the theorem for finite groups too.

We have verified before that groups G satisfying $G = Z_\zeta(G)$ are metacyclic so that the absence of elements of infinite order implies the existence of an ascending chain of subgroups B_ν with the following properties:

(a) $B_0 = 1$;

(b) B_ν is a normal subgroup of $B_{\nu+1}$;

(c) $B_{\nu+1}/B_\nu$ is a cyclic group of order a prime number;

(d) B_ν is for limit-ordinals ν the set of all the elements which are contained in groups B_μ for $\mu < \nu$;

(e) $B_\beta = G$.

* The elements $g^{-i}xg^i$ for integral i form a set containing $p^{n(x)}$ elements. Since $n(1) = 0$, there exists at least one element $y \neq 1$ such that $n(y) = 0$, since the number of elements in K is a power of p.

Since B_0 is the direct product of its primary components, we may assume that it has already been proved that every B_μ for $\mu<\nu$ is the direct product of its primary components $B_{p\mu}$.

CASE 1. $\nu=\rho+1$.

The index of B_ρ in B_ν is a certain prime number p; and B_ρ is the direct product of its primary components $B_{q\rho}$ so that we may represent B_ρ as the direct product of $B_{p\rho}$ and of the direct product H of all the $B_{q\rho}$ for q not equal to p.

Let u be any element generating B_ν modulo B_ρ. The order of u has the form ip^j where i is relatively prime to p and $0<j$. Then u^i generates B_ν modulo B_ρ, so that there exists an element v of order a power of p which generates B_ν modulo B_ρ.

If h is an element in H, then we are going to prove that $hv=vh$. This is certainly true if h is an element in $Z_0(G)$. Hence we may assume that this has been proved for all those elements in H which are contained in groups $Z_\lambda(G)$ for $\lambda<\sigma$. If h is an element in the crosscut of H and $Z_\sigma(G)$ which is not contained in any $Z_\lambda(G)$ for $\lambda<\sigma$, then $c=vhv^{-1}h^{-1}$ is an element in some $Z_\kappa(G)$ for $\kappa<\sigma$. Since H is a characteristic subgroup of B_ρ, and since B_ρ is a normal subgroup of B_ν, H is a normal subgroup of B_ν and c is an element in H. Hence it follows from the induction hypothesis that $cv=vc$, and consequently we find that $v^thv^{-t}=c^th$ for integral t. If t is in particular the order of v, then $c^t=1$. Since the order of v is a power of p, and since the orders of the elements in H are all relatively prime to p, it follows that $c=1$ or $vh=hv$. Since $G=Z_t(G)$, this implies that $hv=vh$ for every h in H.

If x is any element in $B_{p\rho}$, i a positive integer, then

$$(xv^i)^p = xv^ixv^{-i}v^{2i}xv^{-2i}\cdots v^{(p-1)i}xv^{-(p-1)i}v^{pi},$$

and all the factors in the expression on the right are elements in $B_{p\rho}$, so that $(xv^i)^p$ and therefore xv^i itself is an element of order a power of p. The group K, generated by adjoining v to $B_{p\rho}$, is therefore a p-group.

If f is an element in K, h an element in H, then $fh=hf$, since this holds true both for $f=v$ and for f in $B_{p\rho}$. This shows that K contains all the elements of order a power of p in B_ν, and that H contains exactly those elements in B_ν whose orders are prime to p, since B_ν is generated by the elements in K and in H. Hence B_ν is the direct product of K and H, that is, B_ν is the direct product of its primary components.

Case 2. ν is a limit-ordinal.

If u and v are two elements in B_ν whose orders are powers of the same prime number p, then the order of uv is a power of p too, since both u and v are contained in some B_μ for $\mu<\nu$. The elements of order a power of p in B_ν

form therefore a subgroup of B_π; and now it follows that B_π is the direct product of its primary components.

Accordingly every B_π is the direct product of its primary components; and this holds true in particular for $B_\beta = G$.

This proof shows slightly more than we intended to prove. For our argument contains a proof of the following statement.

COROLLARY 3.3. *If all the elements in G are of finite order, if $uv = vu$ for elements u and v in G whose orders are relatively prime, and if G is metacyclic, then G is the direct product of its primary components.*

REMARK. If G is a finite group of order $\prod_p p^{n(p)}$, then it contains a Sylow subgroup S_p of order $p^{n(p)}$; and G is generated by any such system of Sylow representatives. If G satisfies the condition

(5) *$uv = vu$, if the orders of u and v are relatively prime,*

then the subgroup generated by a system of Sylow representatives is their direct product. This shows that *a finite group is a direct product of p-groups if, and only if, it satisfies condition* (5).

It should however be noted that for the proof of this fact we needed the existence of Sylow subgroups, a fact that is comparatively much deeper than the fairly elementary means employed in the proof of Corollary 3.3.

EXAMPLE 3.4. *There exist infinite p-groups whose central is* 1, *though their commutator subgroup is abelian.*

This example shows in particular that it is impossible to omit the solubility in (2) of Theorem 3.2.

Let B be a countable abelian group all of whose elements not equal to 1 are of order the prime number p, so that there exists a basis $b(1), b(2), \cdots, b(i-1), b(i), \cdots$ of B.

As is well known, there exists for every non-negative integer a unique p-adic expansion $\sum_{i=0}^{\infty} c_i p^i$ where $0 \leqq c_i < p$ and where all the c_i—apart from a finite number of exceptions—are 0.

An automorphism $\gamma(j)$ for $0 < j$ of B is defined by

$$b\left(\sum_{i=0}^{\infty} c_i p^i\right)^{\gamma(j)} = b\left(\sum_{i=0}^{\infty} c_{ij} p^i\right)$$

for

$$\begin{aligned} c_{ij} &= c_i, && \text{if} && i \neq j-1, \\ c_{j-1\,j} &= c_{j-1} + 1 && \text{for} && c_{j-1} < p-1, \\ &= 0 && \text{for} && c_{j-1} = p-1. \end{aligned}$$

These automorphisms γ are all of order p and generate a commutative group of automorphisms. Hence there exists a group G which is generated by adjoining to B elements $u(j)$, subject to the relations

$$u(j)^p = 1,\ u(j)u(h) = u(h)u(j), \qquad b(i)u(j) = u(j)b(i)^{\gamma(j)}.$$

B is the commutator subgroup of this group G; and all the elements not equal to 1 in G are of order p.

If U is the subgroup generated by the elements $u(j)$, then U is an abelian subgroup of G, the crosscut of B and U is 1, and every element in G may be represented uniquely in the form bu for b in B and u in U.

If $u \neq 1$, then u does not permute with $b(1)$. If $b \neq 1$, then there exists an integer i so that b is a product of elements $b(j)$ for $0 \leqq j < p^i$; and hence $b \neq b^{\gamma(i+1)}$, so that b does not permute with $u(i+1)$. Thus the central of G is 1.

If b is an element in B, u in U, then $buU = Ubu$ if, and only if, $bU = Ub$. If $b \neq 1$, then there exists an element v in U—as has been pointed out just now—so that $vb \neq bv$. Since the commutator of b and of v is an element not equal to 1 in B, this shows that $bU = Ub$ implies $b = 1$; and we have proved that *the normalizer of U in G is just U.*

Any finite number of elements in G generates a finite subgroup of G. It is a consequence of Theorem 4.1 below that this implies that (G) is satisfied by G, since finite p-groups satisfy (G).

If the subgroup S of G is generated by adjoining a finite number of elements $u(i)$ to B, then S is a normal subgroup of G, so that $S = Z_\sigma(S)$ for some ordinal σ. But the join of these subgroups is G, and $Z(G) = 1$.

3.B. A characteristic property of the groups without elements of infinite order which satisfy condition (G) is given in the following theorem:

THEOREM 3.5. *Suppose that G be a group without elements of infinite order; and denote by $W(G)$ the crosscut of $Z(G)$ and of the iterated commutator subgroups $C_i(G)$. Then condition* (G) *is satisfied by G if, and only if, $G/W(G)$ is the direct product of its primary components, and if these primary components satisfy* (G).

REMARK. It is still an open question whether or not all p-groups satisfy (G), and whether or not there exist groups G without elements of infinite order which satisfy (G), though they are not direct products of p-groups.

Proof. We assume first that (G) is satisfied by the group G; and we derive from this hypothesis that

(3.5.1) *$uv = vu$, if the orders of u and v are relatively prime.*

For, denote by U the subgroup generated by u and v. Since all the elements in G are of finite order, and since (G) is satisfied by G and therefore

by U, it follows from Lemma 1.4 that all the quotient groups $U/C_i(U)$ are finite, and that either $C_i(U)=1$ or $C_{i+1}(U)<C_i(U)$. Hence it follows from Corollary 2.3 that every $U/C_i(U)$ is a direct product of p-groups, so that in particular: $uv\equiv vu$ modulo $C_i(U)$, since the orders of u and v are relatively prime. Hence every $U/C_i(U)$ is abelian, so that $C(U)=C_2(U)$ and consequently $C(U)=1$. Hence U is abelian, and this proves our statement.

Denote now by S the subgroup generated by two elements x and y both of order a power of the same prime number p. The order of xy is a positive number ip^j for $0\leqq j$ and i relatively prime to p. Put $z=(xy)^{p^j}$. Then z is an element of order i. If s is an element of order a power of p, then $sz=zs$ by (3.5.1). If the order of s is relatively prime to p, then it follows from (3.5.1) that $sx=xs$ and $sy=ys$; and this implies $sz=zs$, so that z is an element in $Z(G)$.

It is a consequence of Lemma 1.4 that $S/C_k(S)$ is a finite group. Since this quotient group is generated by two elements of order a power of p, and since it satisfies (G), it follows from Corollary 2.3 that $S/C_k(S)$ is itself a p-group. Hence z is an element in $C_k(S)\leqq C_k(G)$ for every k so that z is an element in $W(G)$. The element xy is therefore modulo $W(G)$ an element of order a power of p; and this proves that $G/W(G)$ is a direct product of p-groups each of which satisfies (G).

Suppose now that the group H is the direct product of its primary components H_p, and that (G) is satisfied by each H_p. Every subgroup S of H is the direct product of its primary components S_p, since S_p is the crosscut of S and H_p. If T is a greatest subgroup of S, then there exists a prime number q so that the primary component $T_q<S_q$. If $p\neq q$, then the join of T and S_p is a subgroup between T and S which is different from S, since it contains only a proper subgroup of S_q, as T is the direct product of its primary components. Hence $S_p\leqq T$ or $S_p=T_p$. If $T_q<K\leqq S_q$, then the join of T and K is a subgroup between T and S which is different from T and which is therefore equal to S. Hence $K=S_q$ so that T_q is a greatest subgroup of S_q. Since S_q satisfies (G), it follows that T_q is a normal subgroup of S_q, so that T is a normal subgroup of S.

Assume now that $G/W(G)$ is the direct product of its primary components, and that these primary components satisfy (G). Then we have proved just now that $G/W(G)$ satisfies (G), and it follows from Lemma 2.6 that G satisfies (G), since $W(G)\leqq Z(G)$.

3.C. If a group G contains both elements of finite and of infinite order, then it may happen that the elements of finite order do not form a subgroup (example: the group generated by two elements u and v, subject to the rela-

tions $u^2=v^2=1$, where uv is of infinite order). Thus it is of interest to find criteria excluding this possibility.

LEMMA 3.6. *If* (G) *is satisfied by the group G, if G is generated by a finite number of elements each of which is of finite order, if there exists a normal subgroup H of G so that H is abelian and so that G/H is a finite group, then G is a finite group.*

Proof. Our statement is certainly true if the index of H in G is 1, since in this case $G=H$, that is, G is an abelian group generated by a finite number of elements each of which is of finite order. Thus we may assume that our statement is true for all the pairs G', H' which satisfy the hypothesis of the lemma, provided the index of H' in G' is smaller than the index of H in G; and we may assume $G\neq H$.

Suppose furthermore that K is a subgroup of H and that K is a normal subgroup of G. Then G/H and $[G/K]/[H/K]$ are essentially the same groups; and all the hypotheses of the lemma concerning G and H are satisfied by G/K and H/K.

It will be convenient to put $G/K=G^*$ and $H/K=H^*$, though the conclusions we are going to arrive at now will later be applied to different subgroups K.

Since G^* is generated by a finite number of elements, and since G^*/H^* is a finite group, it follows from (1.3) that H^* *is generated by a finite number of elements.*

Since G^* satisfies condition (G), it follows from Lemma 1.9 that there exists for every element g in G^* and for every subgroup R of H^*, satisfying $gR=Rg$ and $R\neq 1$, an element $r\neq 1$ in R so that $gr=rg$. Since G^*/H^* is a finite group satisfying (G), it follows from Corollary 2.3 that $G^*/H^*=Z_k(G^*/H^*)$ for some positive integer k; and consequently it follows from Lemma 1.10 that *the crosscut of* $Z(G^*)$ *and* H^* *is different from* 1, *if* $H^*\neq 1$.

Since H is an abelian group, the elements of finite order in H form a subgroup F of H. Our lemma will be proved as soon as we know that $H=F$, since H is generated by a finite number of elements. Thus let us suppose that $F<H$. If K is a subgroup of H so that $H/K\neq 1$ and does not contain elements of finite order not equal to 1, and so that K is a normal subgroup of G, then $F\leqq K$; and the crosscut K'' of the central of G/K and of H/K is different from 1, and modulo K'' there are no elements not 1 of finite order in H/K. If K' is the subgroup between K and H so that $K''=K'/K$, then the rank of H/K' is smaller than the rank of H/K, since both ranks are finite; and the same hypotheses that applied to K apply to K'. Thus we find after a finite number of steps that $F\neq H$ implies the existence of a subgroup N of H with

the following properties: *N is a normal subgroup of G; H/N does not contain any elements not equal to* 1 *of finite order; $N<H$; H/N is part of the central of G/N.*

Denote by Z the subgroup between H and G so that Z/H is the central of G/H. Since G/H is a finite group—which we may assume now to be different from 1—and since G/H satisfies (G), it follows from Corollary 2.3 that $Z/H\neq 1$. The group H/N is part of the central of G/N and is therefore part of the central of Z/N. Since Z/H and $[Z/N]/[H/N]$ are essentially the same, and since the first of these groups is abelian, it follows that Z/N is a group with abelian central quotient group. Since Z/H is a finite abelian group, it follows from the well known properties of groups with abelian central quotient group that all the commutators of elements in Z/N are elements of finite order which are contained in H/N. Since all the elements not 1 in H/N are of infinite order, this proves that *Z/N is an abelian group.*

If $G'=G/N$, $H'=Z/N$, then (G) is satisfied by G', G' is generated by a finite number of elements each of which is of finite order, H' is a normal abelian subgroup of G', and G'/H' is a finite group whose order is smaller than the order of G/H. Hence it follows from the induction hypothesis, stated in the first paragraph of the proof, that G' is a finite group; and this implies in particular that $H'=Z/N$ is a finite group. But this is impossible, since H is between Z and N, and since N has been determined in such a fashion that H/N is an infinite group. Thus we find that the assumption $F\neq H$ leads to a contradiction, that is, $F=H$ is a finite abelian group; and G is a finite group.*

COROLLARY 3.7. *If* (G) *is satisfied by the group G, if G is generated by a finite number of elements each of which is of finite order, then $G/C_i(G)$ is a finite group, $C_i(G)$ is generated by a finite number of elements; and $C_{i-1}(G)\neq 1$ implies $C_i(G)<C_{i-1}(G)$.*

* The classical theorem that a finite group G is a p-group if it satisfies $G=Z_k(G)$ and is generated by elements whose orders are powers of p, may be proved by the same method as follows.

Since the theorem certainly holds true for groups of order 1, we may assume that it holds true for all quotient groups of G whose orders are smaller than the order of G. Thus in particular $G/Z(G)$ is a p-group. If $Z(G)$ itself is a p-group, then our theorem has been proved for G. Otherwise $Z(G)$ is the direct product of two groups P and Q where all the elements in P are of order a power of p, and where the orders of the elements in Q are prime to p. P is a normal subgroup of G, since it is a characteristic subgroup of $Z(G)$. If $P\neq 1$, then the order of G/P is smaller than the order of G, so that G/P would be a p-group and $Q=1$, that is, G itself would be a p-group. Thus assume that $P=1$. Then $Q=Z(G)$ is part of the central of $Z_2(G)=R$, and R/Q is an abelian p-group. Now it follows from the well known properties of groups with abelian central quotient group that the commutators of elements in R are elements in Q whose orders are powers of p. Hence they are all equal to 1, and $Z_2(G)=R$ is an abelian group. As R/Q is a p-group, R is the direct product of Q and a p-group S. If $S=1$, then $Z(G)=Z_2(G)$ and G is an abelian group and as such a p-group. If $S\neq 1$, then G/S is of lower order than G and is therefore a p-group. Since Q is isomorphic with a subgroup of G/S, it follows that $Q=1$ so that G is a p-group in every case.

REMARK. This statement is an improvement upon Lemma 1.4.

Proof. Since our statement is certainly true for $i=0$, we may assume that it has been verified already for $i-1$. Thus $G/C_{i-1}(G)$ is a finite group, $C_{i-1}(G)$ is generated by a finite number of elements, and $C_{i-1}(G)\neq 1$ implies $C_i(G)<C_{i-1}(G)$. If $G'=G/C_i(G)$, $H'=C_{i-1}(G)/C_i(G)$, then G'/H' and $G/C_{i-1}(G)$ are essentially the same finite groups, so that Lemma 3.6 may be applied to G', H'. Hence G' is a finite group. Since therefore $G/C_i(G)$ is a finite group, and since G is generated by a finite number of elements, it follows from (1.3) that $C_i(G)$ is generated by a finite number of elements; and it follows from (1.2) that either $C_i(G)=1$ or $C_{i+1}(G)<C_i(G)$.

COROLLARY 3.8. *If condition* (G) *is satisfied by the group G, and if u, v are elements of relatively prime finite orders in G, then* $uv=vu$.

REMARK. This statement improves upon (3.5.1).

Proof. If W is the group generated by u and v, then it is a consequence of Corollary 3.7 that $W/C_2(W)$ is a finite group. Since $W/C_2(W)$ satisfies (G), and since the orders of u and v are relatively prime, it follows from Corollary 2.3 that $uv\equiv vu$ modulo $C_2(W)$, so that $W/C_2(W)$ is abelian, since this quotient group is generated by u and v. Hence $C(W)=C_2(W)$; and it follows from Corollary 3.7 that $C(W)=C_2(W)=1$, so that W is abelian and in particular $uv=vu$.

3. D. For the following considerations it will be convenient to have a concept intermediate between "metacyclic" and "soluble." A group G shall be termed *weakly soluble* if there exists an ascending chain of subgroups B_ν with the following properties:

(a) $B_0=1$;

(b) B_ν *is a normal subgroup of* G;

(c) $B_{\nu+1}/B_\nu$ *is abelian*;

(d) B_ν *is for limit-ordinals* ν *the set of all the elements contained in subgroups* B_μ *for* $\mu<\nu$;

(e) $B_\tau=G$.

One verifies that soluble groups are weakly soluble, and that weakly soluble groups are metacyclic. If $G\neq 1$, then there is no loss of generality in assuming that $B_\mu<B_\nu$ for $\mu<\nu$.

THEOREM 3.9. *If* (G) *is satisfied by the weakly soluble group G, and if G is generated by a finite number of elements each of which is of finite order, then G is finite.*

REMARK. Finite groups, satisfying (G), are, by Corollary 2.3, soluble groups.

Proof. Since G is weakly soluble, there exists a properly ascending chain of subgroups B_τ, satisfying (a) to (e). The ordinal τ may be represented uniquely in the form $\tau=\sigma+n$ where σ is either 0 or a limit-ordinal, and where n is a non-negative integer. We are now going to prove that $G/B_{\tau-i}$ is a finite group for $0\leqq i\leqq n$. Since this is certainly true for $i=0$, we may suppose that it holds true for $i-1$. If we put $J=G/B_{\tau-i}$, $H=B_{\tau-i+1}/B_{\tau-i}$, then H is abelian, $J/H=G/B_{\tau-i+1}$ is finite, and J is generated by a finite number of elements each of which is of finite order. Hence it follows from Lemma 3.6 that J itself is finite.

Thus in particular G/B_σ is a finite group. Since G is generated by a finite number of elements, it follows from (1.3) that B_σ is generated by a finite number of elements. If σ were a limit-ordinal, then this would imply $B_\sigma=B_\rho$ for some $\rho<\sigma$. This is impossible, since the B_τ form a properly ascending chain of subgroups. Hence $B_\sigma=B_0=1$, so that $G/B_\sigma=G/B_0=G$ is a finite group.

COROLLARY 3.10. *If condition* (G) *is satisfied by the weakly soluble group G, then the subgroup, generated by the elements of finite order in G, is a direct product of p-groups* [*and does not contain elements of infinite order*].

REMARK. It would be sufficient to assume that subgroups of G which are generated by a finite number of elements of finite order satisfy (G) and are weakly soluble. The group generated by two elements u and v, subject to the relations $u^2=v^2=1$, is soluble though uv is an element of infinite order. This shows that condition (G) cannot be omitted. Whether or not the weak solubility is really needed is still an open question.

Proof. If u and v are two elements of finite order, then it follows from the hypothesis and from Theorem 3.9 that they generate a finite group, so that uv is an element of finite order. The elements of finite order in G form therefore a subgroup F without elements of infinite order. It follows furthermore from Corollary 3.8 that $xy=yx$, if x and y are elements in H whose orders are relatively prime. Since G is weakly soluble, F is metacyclic, and it follows now from Corollary 3.3 that F is a direct product of p-groups.

4.A. In this section "properties in the large" will be derived from "properties in the small."

THEOREM 4.1. *Condition* (G) *is a consequence of the following property*:

(G*) *If T is a greatest subgroup of the subgroup S of G, and if S is generated by a finite number of elements, then T is a normal subgroup of S.*

REMARK. This theorem has two important, immediate consequences.

(i) In every group there exists at least one greatest subgroup satisfying condition (G).

(ii) In every group there exists at least one greatest normal subgroup satisfying (G).

Proof. Suppose that S $[<T]$ is a greatest subgroup of the subgroup T of G. Let t be any element in T which is not contained in S, let s be some element in S, and put $c=sts^{-1}t^{-1}$. If c were not contained in S, then T would be generated by adjoining c to S. Hence there would exist a finite number s_i of elements in S so that t is contained in the subgroup V generated by the elements c_1S and s_i. Since the s_i, s are contained in the crosscut U of S and V, and since t is not contained in S, there exists a greatest subgroup W of V which contains $[s, s_i$ and$]$ U, but not t. If $W<M\leqq V$, then t is in M, so that c is in M, since s is in W. Hence M contains s, s_i, c, that is, $M=V$. V is generated by a finite number of elements. W is a greatest subgroup of V. Consequently W is a normal subgroup of V. Since t is in V, and since s is in W, this implies that c is in W, so that t itself belongs to W. This is impossible. Thus it follows that every $tst^{-1}s^{-1}$ for t in T and s in S is an element in S, that is, S is a normal subgroup of T. Hence (G) is satisfied by G.

COROLLARY 4.2. *The group G satisfies* (G) *and has at the same time the property that any finite subset generates a finite subgroup if, and only if, G is the direct product of p-groups with the property that their finite subsets generate finite subgroups.*

REMARK. It is a consequence of this fact that the group constructed as Example 3.4 satisfies condition (G).

Proof. Suppose first that G satisfies (G) and has the property that finite subsets generate finite subgroups. Tnen all the elements in G are of finite order. Let u and v be elements both of order a power of the prime number p. The subgroup generated by u and v is finite and satisfies (G). Hence it follows from Corollary 2.3 that this subgroup is a finite p-group, so that uv is of order a power of p. Now it follows again that G is the direct product of its primary components, and their finite subsets clearly generate finite subgroups.

Suppose conversely that G is the direct product of its primary components G_p and that finite subsets of G_p generate finite subgroups of G_p. Let now R be a finite subset of G. Then there exists clearly a finite set S whose elements are each of them of prime power order so that the subgroup T, generated by S, contains all the elements in R. T is easily verified to be the direct product of finite p-groups. R generates therefore a finite group which, by Corollary 2.3, satisfies (G). Now it follows from Theorem 4.1 that G itself satisfies (G), and this completes the proof.

The following statement is an immediate consequence of Lemma 3.1, (3.5.1) and Corollary 4.2.

THEOREM 4.3. *A metacyclic group without elements of infinite order satisfies condition* (G) *if, and only if, it is the direct product of its primary components.*

Combining this result and Corollary 3.10, the following may be proved.

A weakly soluble group is generated by elements of finite order and satisfies condition (G) *if, and only if, it is a direct product of p-groups.*

4.B. It is possible to derive stronger results, if one restricts the length of the ascending subgroup-chains which exhaust the group (cf. Corollary 2.4).

LEMMA 4.4. *Suppose that* (G) *is satisfied by the group G, and that there exists a finite ascending chain of subgroups B_i with the following properties:*

(i) *B_1 is abelian;*

(ii) *B_i is a normal subgroup of B_{i+1} and B_{i+1}/B_i is a cyclic group;*

(iii) *$B_n = G$.*

Then $G = Z_\zeta(G)$ for some suitable ordinal ζ.

Proof. Suppose that the group G satisfies the hypothesis of the lemma and that $G \neq 1$. Then we may assume that $B_1 \neq 1$.

Clearly $Z(B_1) = B_1 \neq 1$. Hence assume that it has already been proved that $Z(B_i) \neq 1$. Since B_i is a normal subgroup of B_{i+1}, and since $Z(B_i)$ is a characteristic subgroup of B_i, it follows that $Z(B_i)$ is a normal subgroup of B_{i+1}. Since B_{i+1}/B_i is a cyclic group, there exists an element b which generates B_{i+1} modulo B_i. Since $Z(B_i)$ is an abelian group not equal to 1, since $bZ(B_i) = Z(B_i)b$, and since the subgroup generated by b and by $Z(B_i)$ satisfies (G), it follows from Lemma 1.9 that there exists an element $v \neq 1$ in $Z(B_i)$ so that $bv = vb$. Clearly v is an element in $Z(B_{i+1})$, so that this latter subgroup is different from 1. Hence it follows by complete induction that $Z(G) \neq 1$, if $G \neq 1$.

The last result and Lemma 1.11 imply finally that $G = Z_\zeta(G)$.

COROLLARY 4.5. *A group G with abelian commutator subgroup satisfies condition* (G) *if, and only if, its finite subsets generate subgroups S so that $S = Z_\sigma(S)$ for suitable ordinals σ.*

Proof. The sufficiency of the condition is a consequence of Theorems 2.2 and 4.1. If conversely (G) is satisfied by G, and if the subgroup S of G is generated by a finite number of elements, then $C(S)$ is abelian and $S/C(S)$ is generated by a finite number of elements. Hence Lemma 4.4 may be applied to S, and this proves the necessity of the condition.

COROLLARY 4.6. *A metacyclic group of length less than or equal to ω satisfies condition* (G) *if, and only if, its finite subsets generate subgroups S so that $S = Z_\sigma(S)$ for suitable ordinals σ.*

Proof. The sufficiency of the condition is a consequence of Theorems 2.2 and 4.1. If conversely (G) is satisfied by the metacyclic group H of finite length, then it is a consequence of Lemma 4.4 that $H=Z_\eta(H)$ for suitable ordinals η. If G is a metacyclic group of length less than or equal to ω, then the subgroups of G are metacyclic groups of length not exceeding ω, so that finite subsets generate metacyclic groups of finite length; and this proves the necessity of the condition.

4.C. We return now to the analysis of condition (G*) of Theorem 4.1.

THEOREM 4.7. *If the group G is generated by a finite number of elements, then the following two conditions are equivalent*:

(a) *Every greatest subgroup of G is a normal subgroup of G.*

(b) *A set of elements in G generates G if, and only if, it generates G modulo $C(G)$.**

REMARK. Since $G=C(G)$ implies that the identity generates G modulo $C(G)$, it follows that $G\neq 1$, $G=C(G)$ and condition (b) are incompatible.

Proof. Suppose first that the set W of elements in G generates G modulo $C(G)$, but does not generate G. Denote by V the subgroup generated by W, so that $V<G$. Since G is generated by a finite number of elements, it is possible to generate G by adjoining a finite number of elements to V. Let U be a smallest (finite) set of elements in G so that G is generated by V and U. Since $V<G$, it follows that U does not contain the identity and that it contains at least one element u. Denote by T the subgroup of G which is generated by adjoining all the elements not equal to u in U to V. It follows from the minimum-property of U that $T<G$, and that T does not contain u. There exists a greatest subgroup S of G which contains T but not u. If B is a subgroup of G so that $S<B$, then u is an element in B, that is, $B=G$, so that S is a greatest subgroup of G. Since G is generated by adjoining W to $C(G)$, every subgroup of G that contains both W and $C(G)$ is equal to G. Thus a proper subgroup of G cannot contain both W and $C(G)$. S is a proper subgroup of G which contains $W\leqq V\leqq T\leqq S$. Hence $C(G)$ is not part of S. Hence S is not a normal subgroup of G, since otherwise G/S would be a cyclic group and $C(G)\leqq S$. Thus S is a greatest subgroup of G which is not a normal subgroup of G; and (b) is a consequence of (a).

Suppose now that (b) is satisfied by G and that S is a greatest subgroup of G. Then S cannot generate G modulo $C(G)$, since $S<G$. The subgroup T, generated by S and $C(G)$, is therefore a proper subgroup of G, that is,

* This is a generalization of Burnside's minimal basis theorem for groups of order a power of a prime number; cf. for example, Philip Hall, Proceedings of the London Mathematical Society, (2), vol. 36 (1933), pp. 29–95; in particular, pp. 35–36. Also H. Wielandt, Mathematische Zeitschrift, vol. 41 (1936), pp. 281–282.

$S \leqq T \leqq G$; and since S is a greatest subgroup of G, this implies $S=T$ or $C(G)<S$; and S is consequently a normal subgroup of G, so that (a) is a consequence of (b).

COROLLARY 4.8. *The following property of groups G is a necessary and sufficient condition for the validity of condition* (G):

(G**) *If the subgroup S of G is generated by a finite number of elements, and if the subset U of S generates S modulo C(S), then S is generated by U.*

This is a consequence of Theorems 4.1 and 4.7.

4.D. The following property deserves at least a short discussion in this context, since it may be used as a definition for finite nilpotent groups.

(M) *If S is a subgroup of the group G, and if T is a proper subgroup of S, then there exists an element s in S so that sT = Ts though s is not an element in T.*

In other words, proper subgroups are proper subgroups of their normalizers.

LEMMA 4.9. *If* (M) *is satisfied by the group G, then G is metacyclic and satisfies* (G).

REMARK. Example 3.4 shows that metacyclic groups satisfying (G) need not satisfy (M).

Proof. If S is a proper subgroup of the group G, then there exists in G an element g so that $gS=Sg$ and so that S is a proper subgroup of the group T, generated by adjoining g to S. Clearly S is a normal subgroup of T and T/S is a cyclic group; and this implies the metacyclicity of G. If furthermore K is a greatest subgroup of the subgroup H of G, then the normalizer of K in H is different from K and therefore equal to H, so that K is a normal subgroup of H, that is, (G) is satisfied by G.

COROLLARY 4.10. *A group satisfies* (M) *and does not contain elements of infinite order if, and only if, it is a direct product of p-groups, satisfying* (M).

This is a consequence of Lemma 4.9 and Corollary 4.2, since a metacyclic group without elements of infinite order is by Lemma 3.1 finite, when generated by a finite number of elements.

LEMMA 4.11. *If $G=Z_\zeta(G)$ for some ordinal ζ, then* (M) *is satisfied by the group G.*

Proof. Suppose that S is some subgroup of G, and that T is a proper subgroup of S. There exists by Lemma 2.1 an ordinal σ so that $S=Z_\sigma(S)$; and there exists therefore an ordinal τ so that $Z_\tau(S) \leqq T$ but it is not true that $Z_{\tau+1}(S) \leqq T$. There exists therefore an element v in $Z_{\tau+1}(S)$ which is not con-

tained in T. If g is any element in S, then $vgv^{-1}g^{-1}$ is an element in $Z_\tau(S) \leqq T$; and this implies $vT = Tv$, so that (M) is satisfied by the group G.

COROLLARY 4.12. *Properties* (M) *and* (G) *are equivalent properties of soluble as well as of finite groups.*

This is a fairly immediate consequence of Lemmas 4.9 and 4.11, Theorem 2.2 and Corollary 2.3.

Generalizing a concept introduced by H. Wielandt, op. cit. (second paragraph of the second footnote on page 405), we say that the subgroup S of the group G is a *subnormal subgroup of* G whenever there exists an ascending chain of subgroups B_ν with the following properties:

(i) $B_0 = S$;

(ii) *B_ν is a normal subgroup of $B_{\nu+1}$;*

(iii) *B_ν is for limit-ordinals ν the set of all the elements contained in groups B_μ for $\mu < \nu$;*

(iv) $B_\gamma = G$.

THEOREM 4.13. *Each of the following three properties of a group G implies the others:*

(a) *Condition* (M) *is satisfied by the group G.*

(b) *If T is a proper subgroup of G, then T is a proper subgroup of its normalizer in G.*

(c) *Every subgroup of G is a subnormal subgroup of G.*

Proof. Condition (b) is a trivial special case of condition (M), so that (a) implies (b). That (b) implies (c) is a consequence of the fact that every subgroup of G is a normal subgroup of its normalizer in G. Suppose finally that (c) is satisfied by the group G, that S and T are any two subgroups of G and that S is a proper subgroup of T. Since S is a subnormal subgroup of G, there exists an ascending chain of subgroups B_ν, starting with S and satisfying the above conditions (i) to (iv). Denote by T_ν the crosscut of T and B_ν. Then we have $S = T_0$, and there exists some ordinal ρ so that $S = T_\rho < T_{\rho+1}$, since S is a proper subgroup of G, since the B_ν sweep out the whole group G, and since $S = T_\lambda$ whenever λ is a limit-ordinal and $S = T_\mu$ for every $\mu < \lambda$, as follows from (iii). If u is any element in $T_{\rho+1}$, then $uT_\rho u^{-1} = T_\rho$, since u is both an element in T and in $B_{\rho+1}$, so that $uT_\rho u^{-1}$ is both a subgroup of T and, by (ii), of $B_\rho = uB_\rho u^{-1}$. Thus $S = T_\rho$ is both a normal and a proper subgroup of $T_{\rho+1} \leqq T$, so that S is a proper subgroup of its normalizer in T, that is, (a) is a consequence of (c).

We are going to discuss *properties f of group elements* which are subject to the following condition.

(S) *If all the elements of the subgroup U of the group G satisfy property f, if g is an element in G which satisfies property f, and if $gU = Ug$, then every element in the subgroup generated by U and g satisfies property f.*

The two examples of properties satisfying (S) which interest us are:

(a) that of being an element of order a power of the prime number p; and

(b) that of being an element of finite order.

If f is a property of group elements so that (S) is satisfied by f, and if the group unit satisfies f, then there exists always a subgroup all of whose elements satisfy f though it is not a proper subgroup of a subgroup with this property. Such a greatest subgroup of the group G is termed an *f-component of G*.

If U is an f-component of G, and if V is a subgroup of G such that U is a normal subgroup of V, then it follows from (S) that U is exactly the set of all the elements in V which satisfy property f.

The properties f we are going to discuss are subject to another condition.

(J) *If the element g in G satisfies property f, and if α is an automorphism of G, then g^α satisfies property f too.*

If property f satisfies (J), then a characteristic subgroup of G is generated by the elements satisfying f.

LEMMA 4.14. *If the property f of elements in the group G is satisfied by the group unit, if the property f is subject to the conditions* (S) *and* (J), *and if condition* (M) *is satisfied by the group G, then there exists one and only one f-component F of G; and F contains all those elements of G which satisfy property f, so that F is a characteristic subgroup of G.*

Proof. As has been remarked before, there exists always at least one f-component U of G. Denote by U' the normalizer of U in G, and denote by U'' the normalizer of U' in G. Clearly U is a normal subgroup of U', and U' is a normal subgroup of U''. Since U is a normal subgroup of U', it follows from condition (S) that U contains all the elements in U' which satisfy property f; and it follows from (J) that U is a characteristic subgroup of U'. Since U' is a normal subgroup of U'', this implies that U is a normal subgroup of U'', so that U'' is part of the normalizer of U. Hence $U' = U''$. Since however (M) is satisfied by the group G, no proper subgroup of G equals its normalizer in G, so that $U' = G$. Hence U is a normal subgroup of G; and U contains consequently all the elements of property f in G, since U is at the same time an f-component and a normal subgroup of G.

COROLLARY 4.15. *If property* (M) *is satisfied by the group G, then the set $F(G)$ of the elements of finite order in G is a subgroup of G; and $F(G)$ is the direct product of its primary components.*

REMARK. This is not a consequence of Corollary 3.10, since in Corollary 3.10 the group G is supposed to be weakly soluble; and Lemma 4.9 assures only the metacyclicity of the group G. That $F(G)$ is the direct product of its primary components could be derived from Corollary 4.10, since $F(G)$ satisfies (M). But Corollary 4.10 has been proved by reference to an analogous theorem on finite groups whereas the following proof contains a direct proof for the "finite case."

It may be pointed out furthermore that these considerations furnish a simple proof of a theorem proved by H. Wielandt, op. cit. (second paragraph of the second footnote on page 405), since groups satisfying condition (M) are metacyclic, and since metacyclic groups are finite if, and only if, they do not contain elements of infinite order and their length is finite.

Proof. If f is the property of being an element of finite order, then it follows from Lemma 4.14 that there exists one and only one f-component of G; and that this f-component contains all the elements of finite order. Thus $F(G)$ is the uniquely determined f-component of G, and as such $F(G)$ is a subgroup of G. Similarly it follows from Lemma 4.14 that there exists for every prime number p one and only one p-component of G (and of $F(G)$) which contains all the elements of order a power of p in G; and this implies that $F(G)$ is the direct product of its primary components.

4.E. At the end of §2 we mentioned the ascending norm chain of a group G. A more systematic treatment of the condition that $G=N_\nu(G)$ for some suitable ordinal ν may be given here.

One verifies readily that if a group is swept out by its ascending norm chain, then so are its subgroups and quotient groups. It is furthermore not difficult to verify that a group is swept out by its ascending norm chain whenever the group is swept out by its ascending central chain. But it is still an open question whether or not $N(G)\neq 1$ implies $Z(G)\neq 1$. In this respect only the following facts are known.

(a) $Z(G)\leqq N(G)$.

(b) If $Z(G)$ contains elements of infinite order, then $Z(G)=N(G)$.*

(c) If $N(G)$ is hamiltonian, then $Z(G)\neq 1$.†

Since hamiltonian groups are groups with abelian central quotient group, and since the iterated norms are characteristic subgroups, the following is true:

(4.16) *G is weakly soluble, if $G=N_\rho(G)$ for some ordinal ρ.*

It is a consequence of Lemma 2.6 that

* Cf. op. cit. (see the first footnote on page 407) Theorem 3, p. 259.

† R. Baer, Compositio Mathematica, vol. 2 (1935), pp. 241–246.

(4.17) *Condition* (G) *is satisfied by the group* G *whenever* $G=N_\rho(G)$ *for some ordinal* ρ.

As a consequence of these last two statements it may be inferred from Lemma 1.9 and Lemma 1.10 that the following assertion holds true.

(4.18) *If* $G=N_\rho(G)$ *for some ordinal* ρ, *and if* $N(G)$ *is an abelian group of finite rank, then* $Z(G)\neq 1$.

The following fact from the theory of groups with cyclic norm quotient groups will be needed.*

(α) If B is an abelian and normal subgroup of the group G, if G/B is a cyclic group, if $B\leqq N(G)$, if D is the crosscut of B and $Z(G)$, and if $B\neq 1$, then $D\neq 1$ and B/D is a cyclic group.

THEOREM 4.19. *A group* G *which is generated by a finite number of elements satisfies* $G=Z_\zeta(G)$ *for some ordinal* ζ *if, and only if,* $G=N_\rho(G)$ *for some ordinal* ρ.

REMARK. Whether or not the condition that G be generated by a finite number of elements is really needed for the truth of this statement is still an open question. The interest of the above theorem lies in its connection with facts like Corollaries 4.5 and 4.6.

Proof. It is a consequence of facts already related in the course of this discussion, and of Lemma 1.11, that it suffices to prove the following fact.

(4.19.1) *If the group* G *is generated by a finite number of elements, and if* $N(G)$ *is an abelian group of infinite rank, then* $Z(G)\neq 1$.

To prove this statement, let $u_1, \cdots, u_k$ be some finite set of generators of G; and denote by U_i the set of elements y in $N(G)$ so that $yu_i=u_iy$. In applying (α) to the subgroup generated by $N(G)$ and u_i, it follows that $N(G)/U_i$ is a cyclic group. Denote now by V_i the crosscut of the groups U_j for $1\leqq j\leqq i$. Then $V_1=U_1$, so that $N(G)/V_1$ is a cyclic group; and V_{i-1}/V_i is a cyclic group, since V_i is the crosscut of V_{i-1} and U_i. Thus it follows finally that $N(G)/V_k$ is of finite rank. Hence $V_k\neq 1$, since $N(G)$ is of infinite rank. But $V_k\leqq Z(G)$, since G is generated by the elements u_i; and this completes the proof.

(4.20) *If* $G=N_\rho(G)$ *for some ordinal* ρ, *then condition* (M) *is satisfied by the group* G.

Proof. If S is a proper subgroup of G, then there exists an ordinal σ so that $N_\sigma(G)\leqq S$, but so that $N_{\sigma+1}(G)\leqq S$ does not hold. If u is any element in

* Cf. op. cit. (see the first footnote on page 407).

$N_{\sigma+1}(G)$ which is not contained in S, and if s is any element in S, then there exists an integer i and an element t in $N_\sigma(G)$ so that $usu^{-1}=u^it$, and this shows that $uS=Su$. Hence S is a proper subgroup of its normalizer in G; and our statement is a consequence of Theorem 4.13.

(4.21) *If $G=N_\rho(G)$ for some ordinal ρ, then the elements of finite order in G generate a subgroup without elements of infinite order which is the direct product of its primary components.*

This is an immediate consequence of (4.16), (4.17) and of Corollary 3.10.

4.E. If a certain property is not satisfied by the group G, then it may still be possible that G possesses a uniquely determined greatest subgroup with this property. Questions of this type shall occupy us now.

THEOREM 4.22. *In every group G there exists one and only one subgroup $P(G)$ with the following properties*:

(i) *$P(G)$ is a normal subgroup of G;*

(ii) *$P(G)$ is the direct product of its primary components;*

(iii) *if the normal subgroup S of G is the direct product of its primary components, then $S\leqq P(G)$.*

Proof. It clearly suffices to prove the existence of some group $P(G)$.

Suppose first that the normal subgroup S of G is the direct product of its primary components S_p. Then S_p is exactly the set of all the elements of order a power of p in S. Since S is a normal subgroup of G, and since S_p is a characteristic subgroup of S, S_p is a normal subgroup of G. If U is a p-group contained in G, V the subgroup generated by U and S_p, then S_p is a normal subgroup of V, and V/S_p is isomorphic to some subgroup of U. Hence V is a p-group too. This implies in particular that S_p is a subgroup of every greatest p-subgroup of G; or if we denote by $P(G;p)$ the crosscut of all the greatest p-subgroups of G, then $S_p\leqq P(G;p)$.

$P(G;p)$ is by its very construction both a normal subgroup of G and a p-group. The subgroup $P(G)$ of G which is generated by the subgroups $P(G;p)$ is therefore a normal subgroup of G and—as is readily verified—the direct product of the $P(G;p)$. Thus (i) and (ii) are satisfied by $P(G)$. That (iii) is satisfied by $P(G)$ has been shown in the first part of the proof.

It is an immediate consequence of this theorem that a finite group possesses a uniquely determined greatest normal and nilpotent subgroup.

If $F^*(G)$ is the crosscut of all the greatest subgroups without elements of infinite order, then $F^*(G)$ is a characteristic subgroup of G which does not contain elements of infinite order. If S is a normal subgroup of G which does not contain elements of infinite order, then $S\leqq F^*(G)$, since the subgroup

generated by a subgroup without elements of infinite order and by a normal subgroup without elements of infinite order is itself a subgroup without elements of infinite order.

There exist in every group G normal subgroups S which satisfy $S=Z_\sigma(S)$ for some ordinal σ—for example, $S=1$. Thus the subgroup $R(G)$ of G which is generated by all the normal subgroups S of G which satisfy $S=Z_\sigma(S)$ is a well-determined characteristic subgroup of G. Example 3.4 however shows that $R(G)$ need not satisfy $R(G)=Z_\rho[R(G)]$.

(4.23) *$R(G)$ is weakly soluble.*

If T is a proper subgroup of $R(G)$ and a normal subgroup of G, then there exists a normal subgroup S of G so that $S\leqq T$ does not hold and so that $S=Z_\sigma(S)$ for some ordinal σ. Hence there exists a greatest ordinal τ so that $Z_\tau(S)\leqq T$, and clearly $Z_\tau(S)<S$. The subgroup U which is generated by T and $Z_{\tau+1}(S)$ is a normal subgroup of G, contains T as a proper subgroup, and U/T is abelian. Now it is clear how to finish the proof of (4.23).

It is a consequence of (4.23)—at least if G or $R(G)$ satisfies condition (G)—that the elements of finite order in $R(G)$ generate a subgroup which is a direct product of p-groups (Corollary 3.10).

There exist in every group G normal subgroups S which satisfy

(i) $S=Z_\sigma(S)$ for some ordinal σ;

(ii) $Z_{\tau+1}(S)/Z_\tau(S)$ is an abelian group of finite rank.

The subgroup $R^*(G)$ which is generated by all the normal subgroups S of G that satisfy (i) and (ii) is a well-determined characteristic subgroup of G; and clearly $R^*(G)\leqq R(G)$.

(4.24) *$R^*(G)$ is soluble.*

This statement may be proved in essentially the same manner as (4.23).

It is a consequence of (4.24) and Theorem 2.2 that

$$R^*(G) = Z_\alpha[R^*(G)]$$

for some ordinal α, if only G or $R^*(G)$ satisfies condition (G).

5. In this section we are going to prove the equivalence of the following three properties, provided the groups under consideration are subject to suitable restrictions.

(N) *The subgroup S of the subgroup T of the group G is a normal subgroup of T whenever there exists at most one subgroup B so that $S<B<T$.*

(N*) *Every subgroup of G is a normal subgroup of G.*

(F) *If S is some subgroup of G, and if T is a normal subgroup of S, then S/T is not isomorphic to any of the groups $G_{p,e}$.*

Here p is a prime number, e a non-negative integer and $G_{p,e}$ the group generated by two elements u and v, subject to the relations

(a) u and $c=uvu^{-1}v^{-1}$ are both of order p;

(b) $uc=cu$, $vc=cv$, $v^p=c^e$.

Both the central and the commutator subgroup of this group $G_{p,e}$ are generated by the element c of order p. If $p=2$, then all the groups $G_{2,e}$ are isomorphic; if $p\neq 2$, then $G_{p,0}$ and $G_{p,1}$ are not isomorphic and all the other groups $G_{p,e}$ are isomorphic to one of these two groups.

(5.1) $G_{p,e}$ *satisfies* (G), *but not* (N).

Proof. $G_{p,e}$ is a group of order p^3 and every subgroup of order p^2 contains the element c. Thus (G) is satisfied by $G_{p,e}$.

Denote now by S the subgroup generated by the element u. S is of order p. There exists one and only one subgroup of order p^2 which contains u, namely the subgroup generated by u and c. Thus there exists one and only one subgroup, different from U and G, between U and G. On the other hand U is not a normal subgroup of G, so that (N) is not satisfied by $G_{p,e}$.

LEMMA 5.2. *If G is a group with abelian central quotient group, then each of the three properties* (N), (N*) *and* (F) *implies the others.*

Proof. It is obvious that (N) is a consequence of (N*); and it follows from (5.1) that (F) is implied by (N). Assume now that the group G with abelian central quotient group does not satisfy (N*). Then there exists a pair of elements u, v so that uvu^{-1} is not a power of v. Denote by W the subgroup generated by u and v. The central $Z(W)$ of W contains $c=uvu^{-1}v^{-1}$, and c too is not a power of v. There exists furthermore some prime number p so that c is not a power of c^p. (If c is of infinite order, any p will do; if c is of finite order, any prime divisor p of the order of c may be chosen, since $c\neq 1$.) Since c is an element in the central of W, the subgroup D generated by c^p is a normal subgroup of W. Put $W'=W/D$ and $x'=Dx$ for x in W. Then c' is an element of the exact order p which is contained in $Z(W')$, and c' is not a power of v'. It follows now from the well known properties of groups with abelian central quotient group that $W'/Z(W')$ is a direct product of two cyclic groups of order p, that both u'^p and v'^p are in $Z(W')$, and that $Z(W')$ is generated by c', u'^p and v'^p. Since c' is an element of order p which is not a power of v', $Z(W')$ is the direct product of two groups R' and S' so that R' is cyclic and contains c' and so that S' contains v'^p. If S is the subgroup of W which contains D and satisfies $S'=S/D$, then W'/S' and W/S are essentially the same groups. On the other hand W/S is generated by the elements Su and Sv which are just subject to the defining relations of groups $G_{p,e}$. Hence G does not satisfy (F), so that (F) implies (N*), and this completes the proof.

COROLLARY 5.3. *If the group G satisfies $G=Z_h(G)$ for some integer h, then each of the three properties* (N), (N*) *and* (F) *implies the others.*†

Proof. Again it suffices to show that (N*) is a consequence of (F). This is a direct consequence of Lemma 5.2 if $G=Z_2(G)$. Hence assume that $2<h$, that $Z_{h-1}(G)<G$, and that (F) be satisfied by G. We put $G'=G/Z_{h-3}(G)$. Then $1<Z(G')<Z_2(G')<Z_3(G')=G'$ and (F) is satisfied by G' as well as by $G''=G'/Z(G')$. The latter group is a group with abelian central quotient group, and it follows from Lemma 5.2 that (N*) is satisfied by G''. If u' and v' are any two elements in G', then it follows that $c'=u'v'u'^{-1}v'^{-1}$ is modulo $Z(G')$ both a power of u' and of v', so that there exist integers i, j and elements z, w in $Z(G')$ satisfying $u'^iz=c'=v'^jw$. Since u' permutes with z, and v' with w, it follows that c' permutes with both u' and v'. The group W generated by u' and v' is therefore a group with abelian central quotient group, satisfying (F); and it follows from Lemma 5.2 that W satisfies (N*), so that c' is actually a power of both u' and v'. Consequently G' satisfies (N*), and this implies $G'=Z_2(G')$, a contradiction. This completes the proof.

COROLLARY 5.4. *If $C^i(G)\neq 1$ implies $C^{i+1}(G)<C^i(G)$, then each of the three properties* (N), (N*) *and* (F) *implies the others.*

REMARK. Condition (N*) implies $C^2(G)=1$.

Proof. It suffices again to prove that (N*) is a consequence of (F). Thus assume that (F) is satisfied by G. Put $G'=G/C^3(G)$. Then G' satisfies (F) and $G'=Z_3(G')$; and it follows from Corollary 5.3 that G' satisfies (N*). Hence $C^2(G')=1$ and therefore $C^2(G)=C^3(G)$, and this implies from our general hypothesis concerning G that $C^2(G)=C^3(G)=1$, so that $G=G'$ satisfies (N*).

6. In the discussion of the preceding section condition (G) was always a consequence of the other hypothesis involved. This will be different in the present section. Whenever we assume that a group satisfies both the conditions (G) and (F), we shall express this shortly by saying that the group satisfies condition (F-G).

LEMMA 6.1. *If there exists a normal, abelian subgroup H of G so that G/H is abelian, then each of the three properties* (N), (N*) *and* (F-G) *implies the others.*

Proof. We note first that (N*) implies (N), that (N) implies (G) and, by (5.1), (F), so that (F-G) is a consequence of (N). Hence it suffices to prove that (N*) is a consequence of (F-G).

Let g be some element in G, but not in H; and let h be some element in H. Denote by Q the group generated by g and h. Since $C(Q)\leqq C(G)\leqq H$, it follows that $C(Q)$ is abelian, and that $hc=ch$ for every c in $C(Q)$. Since (G) is satisfied

† This is a special case of Theorem 6.5 below.

by G and by Q, it follows therefore from Lemma 1.7 that $C^i(Q)=1$ or $C^{i+1}(Q)<C^i(Q)$. Hence it follows from Corollary 5.4 that (N*) is satisfied by Q. This implies in particular that either $gh=hg$—in case Q is abelian—or that $ghg^{-1}h^{-1}$ is an element of order 2—in case Q is hamiltonian. If in particular h were an element of order 2, then Q would be abelian: $gh=hg$; for if Q were hamiltonian, $Z(Q)$ would contain the elements of order 2 anyway, so that $gh=hg$ and Q would be abelian.

Since $C(G)\leqq H$, and since H is abelian, it follows from the previous argument that $C^2(G)$ contains only elements of order 1 and 2, and that therefore $C^3(G)=1$. Hence it follows from Corollary 5.4 that (N*) is satisfied by the group G.

COROLLARY 6.2. *If $C_i(G)\neq 1$ implies $C_{i+1}(G)<C_i(G)$, then each of the three properties* (N), (N*), (F-G) *implies the others.*

Note that condition (N*) implies $C_2(G)=1$.

Proof. It suffices again to prove that (*N**) is a consequence of (F-G). Then $G/C_3(G)$, $G/C_2(G)$ and $C(G)/C_3(G)$ satisfy condition (F-G) too. Since the commutator subgroups of both the groups $G/C_2(G)$ and $C(G)/C_3(G)$ are abelian, it follows from Lemma 6.1 that these two quotient groups satisfy (N*). If $G/C_2(G)$ is abelian, then $C(G)=C_2(G)$ and therefore $C(G)=C_2(G)=1$, so that $G=G/C_2(G)$ is abelian. If $G/C_2(G)$ is hamiltonian, then its commutator subgroup $C(G)/C_2(G)$ is of order 2. Since $C_2(G)/C_3(G)$ is the commutator subgroup of $C(G)/C_3(G)$, and since the commutator subgroup of a hamiltonian group is not of index 2 in the hamiltonian group, it follows now that $C(G)/C_3(G)$ is abelian, so that $C_2(G)=C_3(G)$ and therefore $C_2(G)=C_3(G)=1$, so that $G=G/C_2(G)$ is hamiltonian, and this completes the proof.

THEOREM 6.3. *If G is a group without elements of infinite order, then each of the three properties* (N), (N*) *and* (F-G) *implies the others.*

Proof. Again it suffices to derive (N*) from (F-G). We prove first the following somewhat more general statement.

(6.3.1) *If the group H is generated by a finite number of elements of finite order, and if H satisfies* (F-G) (*or* (N)), *then H is a finite group all of whose subgroups are normal.*

It is a consequence of our hypothesis that $H/C(H)$ is a finite group, and it follows from (1.3) that $C(H)$ is generated by a finite number of elements. If H is not abelian, then $C(H)\neq 1$ and it follows from (G) and (1.2) that $C_2(H)<C(H)<H$. But $H/C_2(H)$ is not abelian, and it follows from Lemma 6.1 and (F-G) that $H/C_2(H)$ is hamiltonian. This implies that $H/C_2(H)$ is a finite group, since H is generated by a finite number of elements of finite order.

Hence it follows from (1.3) that $C_2(H)$ is generated by a finite number of elements; and if $C_2(H) \neq 1$, then it follows from (G) that $C_3(H) < C_2(H)$. However it has been proved in Corollary 6.2 that (F-G) and $C_3(H) < C_2(H)$ are incompatible, so that $C_2(H) = 1$, and $H = H/C_2(H)$ is a finite hamiltonian group. This completes the proof of (6.3.1).

If u and v are now any two elements of a group G without elements of infinite order which satisfies (F-G), then the elements u and v generate by (6.3.1) a finite subgroup which satisfies (N*). Hence uvu^{-1} is a power of v, so that G itself satisfies (N*).

COROLLARY 6.4. *If the group G satisfies* (F-G) (*or* (N)), *then the set of elements of finite order in G is a subgroup of G which satisfies* (N*).

This is a consequence of (6.3.1) and Theorem 6.3.

THEOREM 6.5. *If there exists an ordinal ζ so that $G = Z_\zeta(G)$, then each of the properties* (N*), (N) *and* (F) *implies the others.*

Note that property (N*) implies $G = Z_2(G)$. Whether or not there exist groups G which satisfy (N) (or (F-G)), though not all their subgroups are normal subgroups, is still an open question.

Proof. It suffices to prove that (N*) is a consequence of (F) and $G = Z_\zeta(G)$. It is a consequence of Theorem 2.2 that $G = Z_\zeta(G)$ implies property (G). In order to prove (N*) it suffices to prove that every $Z_\nu(G)$ satisfies (N*). This will be done by complete (transfinite) induction.

$Z_0(G) = 1$ satisfies (N*). Hence we may assume that every $Z_\mu(G)$ for $\mu < \nu$ satisfies (N*).

CASE 1. $\nu = \rho + 1$.

Then $Z_\rho(G)$ satisfies (N*) so that the group $Z_\nu(G) = G'$ satisfies $C_3(G') = 1$, since $C_2[Z_\rho(G)] = 1$, and since $C(G') \leqq Z_\rho(G)$. Hence it follows from Corollary 6.2 that G' satisfies (N*).

CASE 2. ν is a limit-ordinal.

If u and v are two elements in $Z_\nu(G)$, then they are already contained in some $Z_\mu(G)$ for $\mu < \nu$. Since (N*) is satisfied by $Z_\mu(G)$, this implies that uvu^{-1} is a power of v; and consequently $Z_\nu(G)$ satisfies (N*). This completes the proof. Actually we have proved

COROLLARY 6.6. *If the group G satisfies* (N) (*or* (F)), *then every subgroup $Z_\nu(G)$ satisfies* (N*).

Appendix. In §§2–4 several conditions have been mentioned which were all of them characteristic properties of nilpotent groups, provided the group in question is a finite group. If it is not assumed that the groups are finite,

then those properties are no longer equivalent. It will therefore be convenient to have a list of these properties together with a chart indicating their interrelations.

(A) *If S and T are subgroups of the group G, and if S is a proper subgroup of T, then there exists a normal subgroup N of T so that $S \leqq N < T$.*

(A′) *If S is a proper subgroup of the group G, then there exists a normal subgroup N of G so that $S \leqq N < G$.*

(G) *If S and T are subgroups of the group G, and if S is a greatest subgroup of T, then S is a normal subgroup of T.*

(G*) *If the subgroup S of the group G is generated by a finite number of elements, and if T is a greatest subgroup of S, then T is a normal subgroup of S.*

(G**) *If the subgroup S of the group is generated by a finite number of elements, and if the subset U of S generates S modulo $C(S)$, then S is generated by the set U.*

(G′) *Every greatest subgroup of the group G is a normal subgroup of G.*

(K) *$G = N_\rho(G)$ for some ordinal ρ.*

(K′) *If the subgroup S of the group G is generated by a finite number of elements, then $S = N_\sigma(S)$ for some ordinal σ.*

(L) *$C^\tau(G) = 1$ for some ordinal τ.*

(M) *If S and T are subgroups of the group G, and if S is a proper subgroup of T, then S is a proper subgroup of the normalizer of S in T.*

(M*) *If S is a proper subgroup of the group G, then S is a proper subgroup of the normalizer of S in G.*

(M**) *Every subgroup of G is a subnormal subgroup of the group G.*

(P) *The group G is a direct product of p-groups.*

(P′) *The elements of finite order in the group G generate a subgroup $F(G)$ of G which does not contain elements of infinite order and which is a direct product of p-groups.*

(R) *If u and v are elements of finite order in the group G, and if the orders of u and v are relatively prime, then $uv = vu$.*

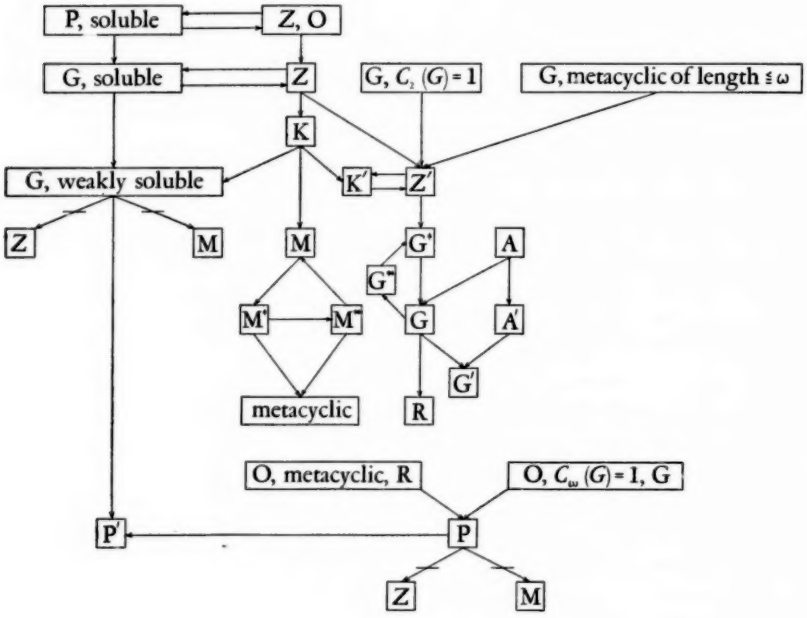
(Z) *$G = Z_\zeta(G)$ for some ordinal ζ.*

(Z′) *If the subgroup S of the group G is generated by a finite number of elements, then $S = Z_\alpha(S)$ for some ordinal α.*

That a finite group, satisfying (L), is a nilpotent group seems to be somewhat accidental if one considers Example 2.5 and the footnote on page 406; and for this reason (L) has not been discussed in this investigation.

In the following chart of inferences we denote by $C_\omega(G)$ the crosscut of the subgroups $C_i(G)$ for integral i; and whenever a group is supposed to consist of elements of finite order only, we indicate this by (O).

Note finally that "arrow crossed out" is used to mean "not implying."



UNIVERSITY OF ILLINOIS,
URBANA, ILL.

NETS AND GROUPS. II*

BY
REINHOLD BAER

In this note we continue the investigation of symmetry-properties of nets, in §8 of a previous paper.† We make free use of the results and definitions of the first five sections of this paper to which reference will be made by prefixing I.

11. In this section we analyze the following property of division-systems with unit.

(S) $x(yz)=1$ *if, and only if,* $(xy)z=1$.

This property may clearly be divided into the following two properties.

(S′) *If* $x(yz)=1$, *then* $(xy)z=1$.

(S″) *If* $(xy)z=1$, *then* $x(yz)=1$.

However the following proposition holds true.

THEOREM 11.1. *If D is a division-system with unit, then each of the properties* (S), (S′), (S″) *implies the others.*

Proof. Clearly it suffices to show that (S″) is a consequence of (S′). Thus let us assume that (S′) holds true, and that $(uv)w=1$. Since D is a division-system with unit, there exists one and only one element r so that $r[u(vw)]=1$. Applying (S′) twice we find that

$$1 = r[u(vw)] = (ru)(vw) = [(ru)v]w.$$

Since there exists one and only one solution of the equation $xw=1$ in D, it follows that $uv=(ru)v$, and that therefore $u=ru$, $r=1$, that is, $u(vw)=1$.

THEOREM 11.2. *Suppose that D is a division-system with unit, and that* $(S<G; r(X))$ *is a canonical representation of D. Then* (S) *is satisfied by D if, and only if,* $r(Z)^{-1}r(Y)^{-1}r(YZ)$ *is an element in S for any two elements Y and Z of D.*

REMARK. $r(YZ)r(Z)^{-1}r(Y)^{-1}$ is always contained in S. Thus our condition amounts to saying that D is not changed, if we consider left-cosets instead of right-cosets.

Proof. Suppose that X, Y, Z are three elements in D. Then $(XY)Z$ is

* Presented to the Society, April 13, 1940; received by the editors November 30, 1939.
† These Transactions, vol. 46 (1939), pp. 110–141.

represented by $r(X)r(Y)r(Z)$; and $X(YZ)$ is represented by $r(X)r(YZ) = r(X)r(Y)r(Z)r(Z)^{-1}r(Y)^{-1}r(YZ)$.

Suppose first that (S) is satisfied by D, and that Y, Z are any two elements in D. Since D is a division-system with unit, there exists one and only one element X in D so that $X(YZ)=1$; and this implies by (S) that both $r(X)r(Y)r(Z)$ and $r(X)r(YZ)$ are elements in S. As S is a subgroup of G, it follows that $r(Z)^{-1}r(Y)^{-1}r(YZ)$ is an element in S and this shows the necessity of our condition.

Suppose conversely that the condition of the theorem is satisfied. If $(XY)Z=1$, then $r(X)r(Y)r(Z)$ is an element in S; and it follows from our condition that $r(X)r(YZ)$ is an element in S, that is, that $X(YZ)=1$. Thus (S'') is satisfied by D; and it follows from Theorem 11.1 that (S) is satisfied by D. This completes the proof.

REMARK. That condition (S) is satisfied by the so-called quasigroups has been proved by Bol.* That the converse does not hold true may be seen from the following example† of a *division-system with unit which satisfies condition* (S), *but which does not satisfy the rule*:

(+) *if* x^{-1} *signifies the solution of* $xy=1$, *then* $(uv)^{-1}=v^{-1}u^{-1}$.

That this last rule is satisfied by all the quasi-groups has been shown by Bol.‡

Denote by H the abelian group of order 4 which is not cyclic; and denote by Z a (cyclic) group of order 2. Then H contains apart from 1 three elements u, v, w of order 2; and Z contains 1 and z. We put $1=f(1,x)=f(x,1)$ for every x in H; $1=f(u,u)=f(v,v)=f(u,v)=f(v,u)$ and $z=f(w,w)=f(u,w)=f(w,u)=f(v,w)=f(w,v)$.

The system D consists of all the (ordered) pairs (h, k) for h in H and k in Z; and multiplication in D is defined by the rule

$$(x, x')(y, y') = (xy, x'y'f(x, y)).$$

It is easily verified that D is a division-system with unit $(1, 1)$, and that D is commutative. We have

$$(1, 1) = (u, 1)^2 = (v, 1)^2 = (w, 1)(w, z),$$

though $(u, 1)(v, 1)=(uv, 1)=(w, 1)\neq(w, 1)^{-1}$, that is, the rule (+) is not satisfied in D.

In order to prove that (S) is satisfied in D, it suffices to consider prod-

* G. Bol, Mathematische Annalen, vol. 114 (1937), pp. 414–431; cf. p. 425.

† This example belongs to a class of systems discovered by Zassenhaus; cf. Bol, op. cit., p. 530.

‡ Cf. Bol, op. cit., p. 425.

ucts of the form $(r, r')[(s, s')(t, t')]$ for $rst=1$. This product is equal to $(1, r's't'f(s, t)f(r, r))$ and we find furthermore that $[(r, r')(s, s')](t, t') = (1, r's't'f(r, s)f(t, t))$. But one verifies that $f(r, r)f(s, t) = f(t, t)f(r, s)$ whenever $rst=1$, since all these products are equal to z, whenever none of the elements r, s, t is equal to 1. Thus (S) is clearly satisfied by D.

If D is a division-system with unit, and if x is an element in D, then we denote by x^{-1} the right inverse of x, that is, x^{-1} is uniquely determined as a solution of the equation $xx^{-1}=1$.

THEOREM 11.3. *If D is a division-system with unit, and if D satisfies condition* (S), *then each of the following properties implies all the others.*

(i) $(xy)(y^{-1}x^{-1})=1$ *for every pair x, y in D.*

(ii) $y(y^{-1}x)=x$ *for every pair x, y in D.*

(iii) $(xy)y^{-1}=x$ *for every pair x, y in D.*

REMARK. The example, discussed just now, shows that these conditions are *not* consequences of (S).

Proof. Suppose first that (i) is satisfied by D. Then it follows from (S) that

$$1 = (xy)(y^{-1}x^{-1}) = x[y(y^{-1}x^{-1})] = [(xy)y^{-1}]x^{-1};$$

and this implies, since D is a division-system, that $x^{-1}=y(y^{-1}x^{-1})$ and $x=(xy)y^{-1}$ so that (ii) and (iii) are both consequences of (i).

If (ii) holds true, then $y(y^{-1}x^{-1})=x^{-1}$ and consequently

$$1 = xx^{-1} = x[y(y^{-1}x^{-1})] = (xy)(y^{-1}x^{-1})$$

by (S) so that (i) is a consequence of (ii); and if (iii) is true, then

$$1 = xx^{-1} = [(xy)y^{-1}]x^{-1} = (xy)(y^{-1}x^{-1})$$

by (S) so that (i) is a consequence of (iii).

The following remark may be of some interest. Condition (iii) is obviously a special case of the following condition:

(iv) $wx=[w(xy)]y^{-1}$ *for any three elements w, x, y, in D.*

Substituting $w=1$, $x=y^{-1}$ we find $y^{-1}=(y^{-1}y)y^{-1}$, that is, $y^{-1}y=1$; and $w=1$ alone gives (iii). Thus we find

$$(wx)y = \{[w(xy)]y^{-1}\}y = \{[w(xy)]y^{-1}\}(y^{-1})^{-1} = w(xy),$$

so that (iv) implies the associative law; that is, *a division-system with unit, satisfying* (iv), *is a group.*

12. In this section we shall need the following special case of the property *R-S.*

PROPERTY *R-S-e*. *If b_1, b_2 are points on $T(e)$, c_1, d_1 on $R(e)$ and a_2, d_2 on $S(e)$, and if $R(a_i)=R(b_i)$, $K(c_i)=R(d_i)$, $S(a_i)=S(d_i)$, $S(b_i)=S(c_i)$, $T(a_1)=T(a_2)$, $T(c_1)=T(c_2)$, then d_1 and d_2 are points on the same T-line.*

This property is illustrated by Fig. 6, I, p. 134.

THEOREM 12.1. *Each of the following properties of a point e in a net N implies the other properties.*

(a) *The system $M(R/S; e)$ satisfies condition* (S) *of* §11.

(b) *The net has Property R-S-e.*

(c) *An anti-isomorphism of $M(R/S; e)$ upon $M(S/R; e)$ is defined in mapping $G(R/S; e)r(R/S; T(e)-X)$ upon $G(S/R; e)r(S/R; T(e)-X)$.*

REMARK. The author is indebted to Dr. Max Zorn for pointing out that the above property (c) is satisfied whenever the system $M(R/S; e)$ is a quasigroup—as has been remarked in §11, condition (S) is satisfied by every quasigroup, though the converse does not hold true. The statement of I, Theorem 8.1, that (2) is equivalent to each of the other conditions is thus erroneous. Regarding its proof the following may be said. The paragraphs preceding I, Fig. 7, p. 135, contain a nearly complete proof of Theorem 12.1, the paragraph containing this figure does not prove anything; and the remaining paragraphs contain a proof of the facts that (3) implies (4), (4) implies (5) and that (5) implies (1). As has been remarked before, it may be verified by a simple computation that (1) implies (3).

The following proof of Theorem 12.1 is slightly different from the one suggested by the discussion in I.

Proof. If Z is any T-line whatsoever, then we shall write for brevity $r(Z)$ instead of $r(R/S; T(e)-Z)$—if X and Y are any two T-lines, then there exists one and only one T-line W so that $e^{r(W)}=e^{r(X)r(Y)}$ and there exists one and only one T-line V so that $e=e^{r(Y)^{-1}r(X)^{-1}r(V)}$, as follows from I, Theorem 4.4. Using the notations of I, Fig. 6, we have $W=T(d_1)$ and $V=T(d_2)$. Thus the net has Property *R-S-e* if, and only if, $W=V$, that is, if, and only if, $e=e^{r(Y)^{-1}r(X)^{-1}r(W)}$, that is, if, and only if, $r(Y)^{-1}r(X)^{-1}r(W)$ is an element in $G(R/S; e)$ for any two T-lines X and Y, that is, by Theorem 11.2, if, and only if, $M(R/S; e)$ satisfies condition (S) of §11; and thus we have proved the equivalence of the properties (a) and (b).

Using again the notations of I, Fig. 6, and applying I, Theorem 4.3, it follows that $r(R/S; T(e)-X)r(R/S; T(e)-Y)$ maps e upon the same point as $r(R/S; T(e)-T(d_1))$ and that $r(S/R; T(e)-Y)r(S/R; T(e)-X)$ maps e upon the same point as $r(S/R; T(e)-T(d_2))$. Thus condition (c) is satisfied if, and only if, $T(d_1)=T(d_2)$, that is, if, and only if, the net has Property *R-S-e*; and thus we have proved that (b) and (c) are equivalent properties.

If D is a division-system with unit, then a *net* $K(D)$ *may be derived from* D in the following fashion. The points of this net $K(D)$ are all the triples (r, s, t) of elements r, s, t in D which satisfy $r(ts)=1$. All the points with the same first coordinate r lie on the same R-line which shall be denoted by r; and the S- and T-lines are defined accordingly. That $K(D)$ is really a net under these definitions is readily derived from the fact that D is a division-system with unit.

THEOREM 12.2.* *Suppose that D is a division-system with unit, that N is the net $K(D)$, just defined, and that $e=(1, 1, 1)$. Then an isomorphism of D upon $M(R/S; e)$ is defined in mapping the element x in D upon the element $G(R/S; e)r(R/S; T(e)-x)$ in $M(R/S; e)$ if, and only if, condition* (S) *of* §11 *is satisfied by D.*

Proof. It is a consequence of I, Theorem 4.3, that $r(R/S; T(e)-x)$ maps the point $p=(r, s, t)$ upon the point $R(p)S\{xR[T(e)S(p)]\}$. As $T(e)=1$, $S(p)=s$, we have $T(e)S(p)=(s', s, 1)$ where s' is the uniquely determined left inverse of s, that is, $s's=1$. Hence $s'=R[T(e)S(p)]$ and $xR[T(e)S(p)]=(s', y, x)$ where y is uniquely determined as the solution of $s'(xy)=1$—the y is the uniquely determined solution of $xy=s$. Thus $S\{xR[T(e)S(p)\}=y$ and the image of p under $r(R/S; T(e)-x)$ is exactly (r, y, v) where v is the uniquely determined solution of $r(vy)=1$ or $vy=r^{-1}$ where r^{-1} is the right inverse to r.

Thus in particular $e=(1, 1, 1)$ is mapped by $r(R/S; T(e)-x)$ upon the point $(1, x^{-1}, x)$.

If u is some other element in D, then it follows furthermore that $(1, x^{-1}, x)$ is mapped by $r(R/S; T(e)-u)$ upon the point $(1, w, w')$ where w' is the left inverse of w and where w is uniquely determined as the solution of the equation $uw=x^{-1}$. The product $r(R/S; T(e)-x)r(R/S; T(e)-u)$ consequently maps the point $e=(1, 1, 1)$ upon the point $(1, w, w')$ where w is determined as the solution of $x(uw)=1$ and w' as the solution of $w'w=1$.

Note now that $r(R/S; T(e)-xu)$ maps the point $e=(1, 1, 1)$ upon the point $(1, (xu)^{-1}, xu)$ so that $r(R/S; T(e)-xu)$ and $r(R/S; T(e)-x)r(R/S; T(e)-u)$ have the same effect upon the point e if, and only if, $w'=xu$. An isomorphism of D upon $M(R/S; e)$ is consequently defined in mapping the element x in D upon the element $G(R/S; e)r(R/S; T(e)-x)$ if, and only if, $x(uw)=1$ implies $(xu)w=1$, that is, if, and only if, D satisfies condition (S′) of §11; and now our theorem is a consequence of Theorem 11.1.

* This theorem shows that just the nets whose coordinates are taken from a division-system with unit, satisfying (S), admit properly of a Kneserian representation; cf. Bol, op. cit., p. 424.

UNIVERSITY OF ILLINOIS,
URBANA, ILL.

REGULAR NORMAL EXTENSIONS OVER COMPLETE FIELDS*

BY

O. F. G. SCHILLING

The local class field theory completely settles all problems pertaining to the abelian extensions of a complete field. All abelian extensions can be described in terms of norm class groups and it can readily be decided which abelian groups can be realized as Galois groups. In this paper we want to attack a more general problem. We consider non-abelian normal extensions of the ground field whose degrees are powers of a prime which is distinct from the characteristic of the residue class field. We construct an infinite normal extension which acts as an universal field. The algebraic approximation of this field yields complete information on the algebraic structure of the field and its Galois group over the ground field. It is shown that the Galois group contains an everywhere dense subgroup which completely suffices to describe the Galois theory. The structure of this group is readily determined. To a certain extent it is a generalization of the fuchsian groups of the classical theory of algebraic functions. The analogy is rather striking. We thus can associate to the given rational prime and the ground field an abstract infinite discrete group which can be considered as a universal covering group with respect to the given prime. All finite groups which can be realized as Galois groups are homomorphic maps of this infinite group. Finally, we associate to certain non-abelian extensions factor groups which are defined in terms of division algebras over the ground field. These factor groups are isomorphic with the respective Galois groups of the fields under consideration. We thus obtain another generalization of local class field theory. In our proofs we make ample use of the ramification theory and local class field theory.

Let k be a field which is complete with respect to a discrete valuation of rank one. Suppose that $\mathfrak{o}$ is the ring of all integers in k and that $\mathfrak{l}=(\lambda)$ is the prime ideal of $\mathfrak{o}$. We shall assume that the residue field $\mathfrak{o}/\mathfrak{l}$ of k is a finite Galois field containing $l^r=q$ elements. Let $p\neq 2$ be a prime such that $q-1\equiv 0 \pmod{p}$. Then k contains the pth roots of unity.† In the sequel we shall sup-

* Presented to the Society, October 28, 1939; received by the editors January 31, 1940. This paper was received by the editors of the Annals of Mathematics August 15, 1939, accepted by them, and later transferred to these Transactions.

† Since k is supposed to be complete, Hensel's irreducibility criterion implies the existence of the $(q-1)$st roots of unity in k. Cf. [4]. The number [4] refers to the reference in the bibliography at the end of this paper.

pose once and for all that k and p are fixed. The structure theory of complete fields yields that k is either a field of $\mathfrak{l}$-adic numbers over the rational l-adic field or a field of formal power series in one variable over the Galois field of q elements (cf. [7, 11, 12.2]).

LEMMA 1. *The maximal abelian extension A of exponent p over k has a Galois group of type (p, p). The field A contains the unramified extension U of degree p over k.*

Proof.† The field A is the join of all radical extensions $k(a^{1/p})$, $a \neq 0$ in k, for k contains the pth roots of unity. Hence, by the general theory of radical extensions, we must investigate the structure of the factor group k^*/k^{*p}, where k^* denotes the multiplicative group of the field k. Let $\{\epsilon\}$ be the group of all units in k which are congruent to 1 (mod $\mathfrak{l}$). We first assert that $\{\epsilon\} = \{\epsilon^p\}$, that is, every unit ϵ is the pth power of a suitable unit $\eta \in \{\epsilon\}$. We shall construct a solution η of $x^p - \epsilon = 0$ by successive approximation. Since $\epsilon \equiv 1 \pmod{\mathfrak{l}}$, we can put $\eta_1 = 1$. Suppose that we already constructed a unit η_i such that $\epsilon \equiv \eta_i^p \pmod{\mathfrak{l}^{i+1}}$. We set $\eta_{i+1} = \eta_i + x_{i+1}\lambda^{i+1}$ where x_i is to be determined in a fixed set of representatives for $\mathfrak{o}/\mathfrak{l}$. We require

$$\epsilon \equiv \eta_{i+1}^p \pmod{\mathfrak{l}^{i+2}}.$$

Consequently, we have $\epsilon \equiv (\eta_i + x_{i+1}\lambda^{i+1})^p \pmod{\mathfrak{l}^{i+2}}$, or $\epsilon \equiv \eta_i^p + p\eta_i^{p-1}x_{i+1}\lambda^{i+1} \pmod{\mathfrak{l}^{i+2}}$; thus

$$(\epsilon - \eta_i^p)(p\eta_i^{p-1})^{-1} \equiv x_{i+1}\lambda^{i+1} \pmod{\mathfrak{l}^{i+2}},$$

for $(p, \mathfrak{l}) = 1$ by assumption. Hence

$$(\epsilon - \eta_i^p)(p\eta_i^{p-1})^{-1}\lambda^{-(i+1)} \equiv x_{i+1} \pmod{\mathfrak{l}},$$

that is, x_{i+1} is uniquely determined by ϵ. Thus k^{*p} contains the group $\{\epsilon\}$. Since k is a complete field, every element $a \in k$ has a unique representation as $\omega^\alpha\lambda^\beta\epsilon$ where ω denotes a fixed primitive $(q-1)$st root of unity and λ a fixed prime element of k. Hence $k^* \pmod{k^{*p}}$ has ω and λ as independent generating representatives. Since $q - 1 \equiv 0 \pmod{p}$, the group k^*/k^{*p} has type (p, p). Consequently, the field A is given as $k(\lambda^{1/p}, \omega^{1/p})$. Obviously, the cyclic unramified field U of degree p over k is contained in A. This completes the proof of the lemma.

Suppose that p^μ is the *highest power* of p which divides $q-1$. Then the maximal abelian extension of exponent p^μ has degree $p^{2\mu}$ and its Galois group has type (p^μ, p^μ). Now let $i > \mu$. We want to find the order of the factor group

† We repeat here Hensel's arguments for sake of completeness. Cf. [5.1].

k^*/k^{*p^i}. Obviously the prime element λ is a representative of order p^i. Consider now the factor group $\{\omega\}/\{\omega^{p^i}\}$. We find for its order

$$[\{\omega\}:\{\omega^{p^i}\}] = [\{\omega\}:\{\omega^p\}] \cdots [\{\omega^\mu\}:\{\omega^{\mu+1}\}][\{\omega^{\mu+1}\}:\{\omega^{\mu+2}\}] \cdots [\{\omega^{p^i-1}\}:\{\omega^{p^i}\}].$$

Hence by the index principle of group theory (considering $\omega \to \omega^p$ as a homomorphism), $[\{\omega\}:\{\omega^{p^\mu}\}] = [\{\omega\}:\{\omega^{p^i}\}] = p^\mu$ for $p^\mu \| q-1$. Thus, we have

LEMMA 2. *The index* $[k^*:k^{*p^i}] = p^{i+\mu}$ *if* $i > \mu$ *and* p^{2i} *if* $i \leqq \mu$.

Suppose now that K is an arbitrary normal extension of k whose degree is a power of p. Let $G = \{x, y, \cdots\}$ be the Galois group of K/k. The elements $xyx^{-1}y^{p-1}$ generate an invarant subgroup G^* of G whose factor group G/G^* has type $(p, p, \cdots, p)$ (cf. [12.1]). Let K^* be the field which corresponds to G^*.

THEOREM 1. *The Galois group of any normal extension K of degree p^n over k can always be generated by two elements.*

Proof. Let G^* be the group $\{xyx^{-1}y^{p-1}\}$. Then the minimal number of generators of G is equal to the number of invariants of G/G^*. This number of invariants is not greater than 2 for $[K^*:k] = [G:G^*] \leqq [A:k] = p^2$ by Lemma 1.

A normal field K over k of degree p^n shall be called a *regular* extension of k.

We now want to construct a *universal field* $N^{(p)}$ over k which contains all regular extensions K. Let $A = A_1$ be the maximal abelian extension of exponent p over k. Since $k \subset A_1$, that is, the pth roots of unity lie in A_1, we can repeat this construction. Let $k \subset A_1 \subset A_2 \subset \cdots \subset A_{i-1} \subset A_i \subset \cdots$ be the infinite tower of relative maximal abelian extensions of exponent p; that is, $[A_i:A_{i-1}] = p^2$, the Galois groups $G(A_i, A_{i-1})$ having type (p, p). A theorem of local class field theory yields that A_i is normal over k, the degree being p^{2i}. Namely, the class group A_{i-1}^{*p} which belongs to A_i/A_{i-1} is left invariant by all elements of $G(A_{i-1}, k)$ (cf. [3]). Consequently, the join $N^{(p)} = \sum A_i$ is an infinite normal extension of k.

THEOREM 2. *The field $N^{(p)}/k$ is universal, that is, it contains all regular fields K/k.*

Proof. The Galois group G of K/k contains a series of normal subgroups $\overline{G}_i$ such that $\overline{G}_i/\overline{G}_{i+1}$ are cyclic groups of order p. Let

$$k \subset K_1 \subset K_2 \subset \cdots \subset K_i \subset \cdots \subset K$$

be the associated chain of subfields (cf. [12.1]). In order to prove our assertion we compare this chain with the defining chain $k \subset A_1 \subset A_2 \subset \cdots$ of $N^{(p)}$.

We have $K_1 \subset A_1$, for K_1 is a cyclic extension of degree p over k. Consider now the cyclic extension K_2 of K_1. Then either $K_2 = A_1$ or $K_2 \neq A_1$. In the second case the join K_2A_1 is a cyclic extension of degree p over A, as a consequence of the Galois theory. Consequently $K_2A_1 \subset A_2$ by construction of A_2. We can continue this process. Thus, ultimately $K \subset A_j$, $\not\subset A_{j-1}$ where the index j is uniquely determined by the given field K.

An immediate consequence of Theorem 2 is the fact that the Galois group G of a regular field K is a homomorphic map $G^{(p)}/S(K)$ of the Galois group $G^{(p)}$ belonging to the universal field $N^{(p)}/k$. We therefore must investigate the structure of $G^{(p)}$ if we want to get information on the various groups G.

Let G_i be the Galois group of an arbitrary but fixed extension A_i over k. Then Theorem 1 implies that G_i can be generated by 2 elements σ_i, τ_i. We shall select σ_i, τ_i such that σ_i generates the inertial group of A_i with respect to k. The group $\{\sigma_i\}$ is an invariant subgroup of G_i and its order is equal to p^i, for A_i contains the unramified field U_{p^i} (of degree p^i over k) as maximal unramified subfield (inertial field) (cf. [4]). Thus

$$G_i/\{\sigma_i\} = G(U_{p^i}, k).$$

The group $G(U_{p^i}, k)$ has order p^i and is generated by an element τ_i^*. Thus, if τ_i denotes a representative of τ_i^* in G_i, we get

$$\{\sigma_i, \tau_i\} = G_i, \qquad \sigma_i^{p^i} = \tau_i^{p^i} = 1, \qquad \tau_i^{-1}\sigma_i\tau_i = \sigma_i^{g_i}$$

where g_i is a prime residue modulo p^i. The group G_{i-1} is a homomorphic map of G_i. Let S_i denote the invariant subgroup of G_i which belongs to A_{i-1}. Then $G_i/S_i \cong G_{i-1}$. It is immediately seen, as a consequence of our selection of the generators σ_i, τ_i, that $S_i = \{\sigma_i^{p^{i-1}}, \tau_i^{p^{i-1}}\}$. Hence we can select the generators σ_{i-1}, τ_{i-1} as the maps of σ_i, τ_i on G_{i-1}. Whence $\tau_{i-1}^{-1}\sigma_{i-1}\tau_{i-1} = \sigma_{i-1}^{g_i} = \sigma_{i-1}^{g_{i-1}}$ and consequently $g_i \equiv g_{i-1} \pmod{p^{i-1}}$.

Next we want to normalize the exponents g_i to an integer g such that $\tau_i^{-1}\sigma_i\tau_i = \sigma_i^g$ for all i. We shall prove this statement by constructing two elements σ, τ in the infinite Galois group of $N^{(p)}/k$ whose homomorphic images in G_i have the required properties.

Suppose that p is an odd rational prime. Denote by V the valuation which is given by p.

LEMMA 3. *Let a be a p-adic integer such that $p^\mu \| a-1$, $\mu \geqq 1$; then there exists a p-adic integer h such that $a^h = 1+p^\mu$. In other words, the p-adic units which are congruent to* 1 (mod p^μ) *form an ideal cyclic group which is generated by any* $a \equiv 1\ (p^\mu)$, $a \not\equiv 1\ (p^{\mu+1})$.

Proof. The exponent h for which $a^h = 1+p^\mu$ will be constructed by suc-

cessive p-adic approximation. Let $a=1+bp^\mu$ where b is a p-adic integer of value 0. Suppose that we already found integers $h_1, \cdots, h_j$ for which

$$h_\nu \equiv h_{\nu-1} \pmod{p^{\nu-1}}, \qquad \nu = 1, 2, \cdots, j, \tag{1}$$

and

$$a^{h_\nu} \equiv 1 + p^\mu \pmod{p^{\mu+\nu}}, \qquad \nu = 1, 2, \cdots, j. \tag{2}$$

The first number h_1 is easily determined. Namely, we require

$$a^{h_1} = (1 + bp^\mu)^{h_1} = 1 + h_1 bp^\mu + p^{2\mu}c, \qquad V(c) \geqq 0.$$

Since $b \not\equiv 0 \pmod{p}$, we can determine h_1 by the congruence $h_1 b \equiv 1 \pmod{p}$. Consequently,

$$a^{h_1} \equiv 1 + p^\mu \pmod{p^{\mu+1}}.$$

In the general case we propose to find an integer h_{j+1} as $h_j + r_j p^j$ where r_j has to be determined. Such a number h_{j+1} surely satisfies condition (1). In order to show that condition (2) can be realized we proceed as follows. We must have

$$a^{h_{j+1}} = a^{h_j} a^{p^j r_j} = (1 + p^\mu + sp^{\mu+j})(1 + bp^\mu)^{r_j p^j}$$

where s is a p-adic integer by induction. Next

$$a^{r_j p^j} = (1 + bp^\mu)^{r_j p^j} = 1 + r_j bp^{\mu+j} + \sum_{\kappa=2}^{r_j p^j} C_{r_j p^j, \kappa} (bp^\mu)^\kappa.$$

We now show that the value of the sum on the right hand exceeds $\mu+j$. The binomial coefficients $C_{r_j p^j, \kappa} = B_j$ have the form $p^j r_j (\kappa!)^{-1} d$ where $V(d) \geqq 0$. Hence

$$V(B_j) \geqq j - V(\kappa!) \geqq j - \kappa(p-1)^{-1},$$

for $V(\kappa!) = (\kappa - s_\kappa)(p-1)^{-1}$ where $s_\kappa \geqq 0$ (cf. [5.2]). Thus

$$\begin{aligned} V(B_j (bp^\mu)^\kappa) &\geqq \mu\kappa + j - \kappa(p-1)^{-1} = \mu + j + \mu(\kappa - 1) - \kappa(p-1)^{-1} \\ &\geqq \mu + j + \mu(\kappa - 1) - \kappa/2 \geqq \mu + j + (\kappa - 1) - \kappa/2 > \mu + j, \end{aligned}$$

provided that $\mu \geqq 1$, $p \geqq 3$, $\kappa \geqq 3$. If $\kappa = 2$, then

$$B_j (bp^\mu)^2 = r_j p^j (r_j p^j - 1) 2^{-1} b^2 p^{2\mu} = p^{2\mu+j} d,$$

where d is a p-adic integer. Consequently,

$$V(B_j (bp^\mu)^2) = 2\mu + j \geqq \mu + j + 1.$$

Hence, in general, $a^{p^j r_j} = 1 + br_j p^{\mu+j} + tp^{\mu+j+1}$, where $V(t) \geqq 0$. Returning to $a^{h_{j+1}}$ we get

$$a^{h_{j+1}} = (1 + p^{\mu} + sp^{\mu+j})(1 + br_j p^{\mu+j} + tp^{\mu+j+1})$$
$$\equiv 1 + p^{\mu} + (s + br_j)p^{\mu+j} \pmod{p^{\mu+j+1}}.$$

Since $b \not\equiv 0 \pmod{p}$, we can determine r_j as a solution of $s + br_j \equiv 0 \pmod{p}$. Hence the induction is completed. The numbers h_j form a convergent p-adic sequence. Let $\lim_{j\to\infty} h_j = h$. Then

$$a^h = a^{\lim h_j} = \lim_{j\to\infty} a^{h_j} = 1 + p^{\mu}.$$

LEMMA 4. *The Galois group G_i of A_i/k is given by the relations $\sigma_i^{p^i} = \tau_i^{p^i} = 1$, $\tau_i^{-1}\sigma_i\tau_i = \sigma_i^{1+p^{\mu}}$ where $a - 1 = p^{\mu} \cdot r$, $(r, p) = 1$.*

Proof. Let $G^{(p)}$ be the infinite Galois group of $N^{(p)}/k$. As usual we define $G^{(p)}$ as the group of vectors $(\rho_1, \cdots, \rho_i, \cdots)$ where (cf. [6])

$$\rho_i \in G_i, \quad \rho_i \pmod{S_i} = \rho_{i-1}, \qquad i = 1, 2, \cdots.$$

Application of the ramification theory of infinite normal extensions yields that $G^{(p)}$ contains two elements σ, τ^* having the following properties. The element σ is a generator of the inertial group of $N^{(p)}/k$, that is, the elements of the inertial group are powers σ^c where the c's are p-adic integers. The group of the inertial field of $N^{(p)}/k$ is ideal cyclic and generated by a residue class $\tau^* \pmod{\{\sigma\}}$ where $\tau^* \in G^{(p)}$ (cf. [6]). Since the inertial group is an invariant subgroup of $G^{(p)}$, we have $\tau^{*-1}\sigma^*\tau^* = \sigma^a$ where a is a p-adic integer. Since $\lim_{i\to\infty} G_i = N^{(p)}$, we have $\sigma = (\sigma_1, \cdots, \sigma_i, \cdots)$, $\sigma_i \in G_i$, and $\tau^* = (\tau_1^*, \cdots, \tau_i^*, \cdots)$, $\tau_i^* \in G_i$. Moreover, $\tau_i^{*-1}\sigma_i\tau_i^* = \sigma_i^{a_i}$ where $a_i \equiv a \pmod{p^i}$ and $\sigma_i^{p^i} = \tau_i^{*p^i} = 1$ according to the selection of the elements σ, τ^* in $G^{(p)}$.

We now want to apply Lemma 3 in order to normalize the exponents g_i. Consider for this purpose the norm groups $N_i A_i^*$ which belong to the various fields A_i/k; N_i denotes the norm taken from A_i to k. Observing that N_i can be split up into the various relative norms from A_j to A_{j-1} and that A_j is the maximal abelian extension of type (p, p) over A_{j-1}, we get $N_i A_i^* = k^{*p^i}$. Hence

$$[k^* : N_i A_i^*] = p^{2i} \quad \text{if} \quad i \leqq \mu$$
$$= p^{\mu+i} \quad \text{if} \quad i > \mu.$$

A theorem of local class field theory yields† that

(i) A_i is an abelian extension of type (p^i, p^i) over k if $i \leqq \mu$,

(ii) A_i is a non-abelian extension of degree p^{2i} over k if $i > \mu$ and the group $N_i A_i^*$ belongs to the maximal abelian subfield $A_\mu U_{p^i}$ of A_i.

Since G_μ is the *last* Galois group in the approximation of $G^{(p)}$ which is abelian, we must have $a \equiv 1 \pmod{p^\mu}$, $a \not\equiv 1 \pmod{p^{\mu+1}}$. Hence there exists,

† [3, 10]. We mean the "Abgrenzungssatz" of local class field theory.

by Lemma 3, a p-adic integer h such that $a^h = 1 + p^\mu$. Consequently, we can change τ^* to a new element $\tau = \tau^{*h}$ in $G^{(p)}$ such that

$$\tau^{-1}\sigma\tau = \sigma^{1+p^\mu}.$$

Namely, $\tau^{*-1}\sigma\tau^* = \sigma^a$ implies $\tau^{*-h}\sigma\tau^{*h} = \sigma^{a^h} = \sigma^{1+p^\mu}$. Hence, also $\tau_i^{-1}\sigma_i\tau_i = \sigma_i^{1+p^\mu}$, $i = 1, 2, \cdots$.

LEMMA 5. *Let L be an infinite normal extension of the abstract field R and let Γ be the (topologized) Galois group of L/R. Suppose that Δ is a subgroup of Γ. Then Δ is everywhere dense in Γ if and only if whenever W/R is a finite normal subfield of L every automorphism of W/R can be extended to an automorphism of Δ.*†

Proof. We first note that a group Δ as described in the lemma is sufficient to describe the Galois theory of the finite extensions W/R which lie in L. An alternate formulation of the conditions of the lemma is this:

If $L \supset W \supset R$, $[W:R]$ finite, and if $\Delta(W)$ is the subgroup of Δ leaving the elements of W fixed, then $\Delta/\Delta(W)$ is isomorphic to $G(W, R)$.

Suppose now that Δ is an everywhere dense subgroup of Γ. We want to prove that the Galois theory for finite extensions $R \subset W \subset L$ can be described in terms of Δ. Let ρ be an arbitrary automorphism of $G(W, R)$. We can extend ρ to an automorphism ρ' of Γ. Since Δ is everywhere dense in Γ and since the automorphisms leaving W elementwise fixed constitute a neighborhood N of the unit in Γ, there exists an element δ in Δ such that $\delta\rho'^{-1}$ lies in N. Thus, δ lies in $N\rho'$ or δ induces ρ on W. Conversely, the possibility of describing the Galois theory for finite extensions W/R by means of a subgroup Δ of Γ implies that Δ is everywhere dense in Γ. Let ρ' be any automorphism of Γ and let $N(W)$ be any neighborhood of the unit in Γ, that is, $N(W)$ is defined by a finite subfield W/R of L. Then there exists, by hypothesis, an element δ in Δ such that δ agrees with ρ' on W, that is, $\delta\rho'^{-1}$ induces the identity on W. Hence $\delta\rho'^{-1}$ is an element of $N(W)$. This means that Δ is an everywhere dense subgroup of Γ.

THEOREM 3. *The group $\{\sigma, \tau; \tau^{-1}\sigma\tau = \sigma^{1+p^\mu}\} = F^{(p)}$ is an everywhere dense subgroup of $G^{(p)}$. The Galois theory for finite subfields of $N^{(p)}/k$ can be described in terms of the subgroup $F^{(p)}$.*

Proof. By Lemma 5 it suffices to prove the first part of the theorem. The topology of $G^{(p)}$ is given by the chain of homomorphisms $G_i \to G_{i-1}$, explicitly

$$\sigma_{i-1} = \sigma_i \pmod{\{\sigma_i^{p^{i-1}}, \tau_i^{p^{i-1}}\}}, \qquad \tau_{i-1} = \tau_i \pmod{\{\sigma_i^{p^{i-1}}, \tau_i^{p^{i-1}}\}}.$$

† The author wants to thank Professor Saunders MacLane for valuable suggestions in the proofs of Lemmas 3, 5.

In other words, we get an isomorphic representation of G_{i-1} by reducing the exponents of the elements $\rho_i = \sigma_i^{\alpha}\tau_i^{\beta}$ in G_i modulo p^{i-1} and changing the subscript i to $i-1$. Since $\sigma_i^{p^i} = \tau_i^{p^i} = 1$, the exponents α, β of the element $\rho_i \in G_i$ are p-adic integers reduced modulo p^i. Next we remark that $\{\sigma^{p^i}, \tau^{p^i}\}$ is an invariant subgroup F_i of $F^{(p)}$. Namely, we have

$$\begin{aligned}\tau^{-1}\sigma^{p^i}\tau &= \tau^{-1}\sigma\tau\tau^{-1}\sigma\tau \cdots \tau^{-1}\sigma\tau \\ &= \sigma^{1+p^{\mu}}\sigma^{1+p^{\mu}} \cdots \sigma^{1+p^{\mu}} \\ &= \sigma^{(1+p^{\mu})p^i} = (\sigma^{p^i})^{1+p^{\mu}} \in \{\sigma^{p^i}, \tau^{p^i}\}.\end{aligned}$$

Next $\sigma\tau\sigma^{-1} = \tau\sigma^{p^{\mu}}$; hence

$$\sigma\tau^{p^i}\sigma^{-1} = \sigma\tau\sigma^{-1}\sigma\tau\sigma^{-1} \cdots \sigma\tau\sigma^{-1} = \tau\sigma^{p^{\mu}}\tau\sigma^{p^{\mu}} \cdots \tau\sigma^{p^{\mu}} = \tau^{p^i}(\sigma^{p^i})^x$$

where $x = [(1+p^{\mu})^{p^i} - 1]p^{-i}$. Thus $F^{(p)}/F_i \cong G_i$. In other words, we obtain G_i from $F^{(p)}$ by reducing the exponents α, β of $\sigma^{\alpha}\tau^{\beta}$ (α, β are rational integers) modulo p^i and attaching the subscript i to σ, τ. Consequently, every element $\rho = (\rho_1, \cdots, \rho_i, \cdots)$ of $G^{(p)}$ can be approximated by a sequence $\rho^{(i)} = \sigma^{\alpha(i)}\tau^{\beta(i)}$, $i = 1, 2, \cdots$, where the $\alpha(i)$, $\beta(i)$ are integers such that

$$\rho_i = \rho^{(i)} \pmod{S^{(i)}} = \sigma^{\alpha(i)}\tau^{\beta(i)} \pmod{F^{(i)}}$$

for every i, where $G^{(p)}/S^{(i)} \cong G_i$. This proves that $F^{(p)}$ is an everywhere dense subgroup of $G^{(p)}$. We remark that the closure $G^{(p)}$ of $F^{(p)}$ is obtained by admitting for the exponents α, β of $\rho = \sigma^{\alpha}\tau^{\beta}$ arbitrary p-adic integers instead of rational integers. The closure of F_i in $G^{(p)}$ is equal to $S^{(i)}$ and $S^{(i)} \cap F^{(p)} = F_i$. This follows immediately from the imbedding of $F^{(p)}$ in $G^{(p)}$.

THEOREM 4. *A finite group G of order p^n can be realized as the Galois group of a reguar extension K/k if and only if it is a homomorphic map of the group $F^{(p)} = \{\sigma, \tau; \tau^{-1}\sigma\tau = \sigma^{1+p^{\mu}}\}$.*

Proof. Let K be an arbitrary regular extension of k. Then $K \subseteqq A_i$, i sufficiently large, by Theorem 2. Then, by Theorem 3, the Galois group G of K is a homomorphic map of $F^{(p)}$. Conversely, suppose that $G \cong F^{(p)}/T$. Then, by the second half of Theorem 3, there exists at least one field $K \supset k$ whose Galois group is isomorphic with G. The field K belongs to T. However, T need not be uniquely determined by G, for the homomorphism $F^{(p)} \to G$ can, in general, be realized by various invariant subgroups T. The number of groups T corresponding to a given group G is finite. Namely, the argument which we used to prove that $K \subset A_i$ shows that i is bounded by n, where $[K:k] = p^n$. Hence $G \cong G(K, k) = G_i/G(A_i, K) = (F^{(p)}/F_i)/(T/F_i)$. We remark that $T^{(1)} \neq T^{(2)}$ implies that the closures in $G^{(p)}$ of T_1, T_2, respectively, are also distinct provided that $T^{(1)}$ and $T^{(2)}$ have finite indices under $F^{(p)}$.

It is relatively easy to determine the exact number of regular fields K with prescribed Galois group of small order. We want to discuss briefly the simplest case. Suppose that $\mu=1$, that is, $q-1=p\cdot r$, $(r, p)=1$. Then A_2 is a non-abelian extension of degree p^4 over k. Its Galois group is given by the relations

$$\Sigma^{p^2}=\Gamma^{p^2}=1, \qquad \Gamma^{-1}\Sigma\Gamma=\Sigma^{1+p}.$$

Now let G be an arbitrary group of order p^3. Since the commutative cases are already settled by local class field theory, we investigate the non-abelian cases. There are two distinct types of non-abelian groups of order p^3:

(α) $\sigma^{p^2}=\tau^p=1$, $\tau^{-1}\sigma\tau=\sigma^{1+p}$ and

(β) $\sigma^p=\tau^p=\rho^p=1$, $\rho=\sigma\tau\sigma^{-1}\tau^{-1}$, $\rho^{-1}\sigma\rho=\sigma$, $\tau^{-1}\sigma\tau=\sigma$, $\rho^{-1}\tau\rho=\tau\sigma$ (cf. [2]).

A simple computation yields that both groups are homomorphic maps of $G_2=\{\Sigma, \Gamma\}$. Hence every type has at least one realization as the Galois group $G(K)$ of a regular field $K\subset A_2$. Moreover, every K which belongs to either type (α) or type (β) must be contained in A_2. Since these fields K are non-abelian extensions of k, they all must contain the field A_1. This field A_1 belongs [12.1] to the group $G_2^*=\{\Sigma\Gamma\Sigma^{-1}\Gamma^{p-1}\}$ of G_2. Since $K\supset A_1$, we also have $G/G^*\cong G_2/G_2^*$ where G^* is defined in the same fashion as G_2^*. Moreover, $G_2/S'(K)=G(K)$. Hence, by the homomorphism principle of group theory, $S'(K)\subset G_2^*$. Thus, the fields K are found among the cyclic extensions of degree p over A_1. There are $p+1$ such extensions for G_2^* has $p+1$ invariant subgroups of order p, as follows from the structure of G_2. Namely, $G_2^*=\{G_2'; X^p, X\in G_2\}$ where G_2' is the commutator group of G_2 (cf. [12.1]). Consequently $G_2^*=\{\Sigma^p, \Gamma^p\}$ for $\Sigma^{-1}\Gamma^{-1}\Sigma\Gamma=\Gamma^p\in\{\Sigma^p, \Gamma^p\}$. Moreover $\Gamma^{-p}\Sigma\Gamma^p=\Sigma$ and $\Gamma^{-1}\Sigma^p\Gamma=\Sigma^p$.

Consider now the invariant subgroup $\{\Sigma^p\}$ of G_2. Its factor group $G_2/\{\Sigma^p\}$ is abelian and has type (p, p^2). The associated field, belonging to G_2', is $U_{p^2}k(\lambda^{1/p})$.

The factor group $G_2/\{\Gamma^p\}$ is non-abelian. It has type (α).

The $p-1$ different subgroups given by $\{\Sigma^{ip}\Gamma^{ip}\}$ have factor groups which belong to type (β). This is easily verified.

Combining these results we can state the following theorem.

THEOREM 5. *If $p\|q-1$, then k has exactly one regular extension of type (α) and exactly ($p-1$) distinct regular extensions of type (β).*

The infinite discrete group $F^{(p)}$ can be considered as a generalization of the fuchsian groups of the classical theory of algebraic functions to l-adic number fields. The analogy is rather striking. We may interpret it as follows. The substitution $\tau\in F^{(p)}$ corresponds to the unramified extensions, that is,

to the fundamental group of the classical case. The substitution $\sigma \in F^{(p)}$ corresponds to the substitution of signature $\lim_{i\to\infty} p^{-i}=p^{-\infty}$ around the given branch point of the underlying Riemann surface. Moreover, the commutator group $F^{(p)\prime}$ of $F^{(p)}$ is given by $\tau^{-1}\sigma\tau\sigma^{-1}=\sigma^{p^\mu}$ as is readily verified. The factor group $F^{(p)}/F^{(p)\prime}$ has type (p^μ, p^∞). Using the ramification theory of infinite normal extensions of an l-adic number field, one sees that the subfield L belonging to $F^{(p)\prime}$ is given as $k(\lambda^{1/p^\mu})U_\infty^{(p)}$ where $U_\infty^{(p)}$ denotes the join of all unramified extensions of degree p^i $(i=1, 2, \cdots)$ over k (cf. [6]). We remark that $F^{(p)}/F^{(p)\prime}=\{\tau, \sigma; \sigma^{p^\mu}=1\}$ is the generalization of the Betti group belonging to a given branch point with signature $1/p^\mu$.

The equivalent to the multiplicative functions belonging to a fuchsian group is given by the multiplicative group $N^{(p)*}$ of the universal field. The field $N^{(p)}$ contains all $p^{i+\mu}$th roots of unity ω_i and all radicals $\lambda^{1/p^i}=\lambda_i$, $i=1, 2, \cdots$. Since k is supposed to contain the pth roots of unity, we have $U_\infty^{(p)}=k(\omega_1, \omega_2, \cdots, \omega_i, \cdots)$. Moreover, the construction of $N^{(p)}$ as $\sum A_i$ implies that $\lambda_i \in N^{(p)}$. Hence

$$N^{(p)} = k(\omega_1, \cdots, \omega_i, \cdots; \lambda_1, \cdots, \lambda_i, \cdots).$$

The groups $\{\omega_i, \lambda_i\}$ which are generated by the elements ω_i, λ_i—for a fixed λ in k—form a multiplicative subgroup $M^{(p)}$ of $N^{(p)*}$. The complete closure $\overline{N^{(p)}}$ of $N^{(p)}$ consists of all sums $\Phi\Psi\sum_{j=0}^{\infty}\Phi_j\Psi_j=P$ where $\Phi, \Phi_j \in \{\omega_i\}, \Psi, \Psi_j \in \{\lambda_i\}$ such that $V(\Psi_j)\to\infty$ where V denotes the valuation of $N^{(p)}$;† $V(\Psi_0)=0$, $\Phi_0\neq 0$. Thus $N^{(p)}$ consists of all elements of $\overline{N^{(p)}}$ which are algebraic over k.

Now let K be an arbitrary regular finite extension of k. The structure theory of complete fields yields that K is complete and that it is uniquely determined by the root of unity $\omega(K)$ of highest order which is contained in K and a prime element $\lambda(K)$ of K:

$$K = \left(\omega(K)^a\lambda(K)^b\sum_{j=0}^{\infty}\omega(K)^{c_j}\lambda(K)^j\right),$$

the c_j residues modulo the order of $\omega(K)$ (cf. [7, 11, 12.2]).

According to Theorem 2 we have $K \subseteq A_i$, where i is sufficiently large. Since $A_i=k(\lambda_i, \omega_i)$, we have

$$\omega(K) = \omega_i^{\alpha(K)}, \qquad \lambda(K) = \omega_i^{\beta(K)}\lambda_i^{\delta(K)}\cdot\epsilon_i$$

where ϵ_i is a unit of A_i which is congruent to 1 (mod (λ_i)).

Let $\{\rho\}$ be the Galois group of A_i/K. Then

$$\lambda(K) = \lambda(K)^\rho = (\omega_i^{\beta(K)}\lambda_i^{\delta(K)}\epsilon_i)^\rho = (\omega_i^{\beta(K)}\lambda_i^{\delta(K)})^\rho\epsilon_i^\rho.$$

† [9]. By Hensel's criterion applied to relatively complete fields.

Hence $(\omega_i^{\beta(K)}\lambda_i^{\delta(K)})^{\rho-1}=\epsilon_i^{1-\rho}$ for all ρ. The unit $\epsilon_i^{1-\rho}$ lies in $\{\epsilon_i\}$; moreover $\{\omega_i, \lambda_i\} \cap \{\epsilon_i\}=1$. Thus, $\epsilon_i^{1-\rho}=1$ for all ρ. For ρ effects a multiplication of λ_i by a root of unity according to the definition of λ_i as a radical. Hence $\epsilon_i \in K$. Thus, $\lambda(K)\cdot\epsilon_i^{-1}=\lambda'(K)=\omega_i^{\beta(K)}\lambda_i^{\delta(K)}$. In other words, the prime element $\lambda(K)$ of K can be normalized in such a way that it lies in $\{\lambda_i, \omega_i\}$.

The group $F^{(p)}$ acts as operator group on $M^{(p)}$. Arranging the elements of $M^{(p)}$ as follows: $\{\{\omega_1, \lambda_1\}, \cdots, \{\omega_i, \lambda_i\}\}$, and observing that $F^{(p)}$ is everywhere dense in $G^{(p)}$, it follows that this representation of $F^{(p)}$ as operator group is an isomorphic one. Let S be an arbitrary normal subgroup of finite index under $F^{(p)}$. Denote by $M(S)$ the subgroup of all elements in $M^{(p)}$ which are left invariant by S. Since S contains some S_i belonging to a field A_i, it follows that $M(S) \subseteqq \{\omega_i, \lambda_i\}$. We already noticed before that the field of S is a uniquely determined subfield $K(S)$ of A_i. Hence we can apply the result concerning the normalization of the prime element of $K(S)$. It follows that $M(S)$ is equal to $\{\omega_i, \lambda_i\} \cap K(S)$. In order to determine the maximal root of unity $\omega(S)$ and the normalized prime element $\lambda(S)$ of $K(S)$ we proceed as follows. Since $M(S)$ is a finite group contained in $\{\omega_i, \lambda_i\}$, it has a minimal base of at most 2 elements. Let $\lambda(S)$ be an element of $M(S)$ such that $V(\lambda(S))$ is minimal. Then take for $\omega(S)$ an arbitrary other element of value 0 which has maximal order. The elements $\lambda(S)$, $\omega(S)$, thus selected, obviously have the required properties. These results which are implied by the fact that we have to deal with (discrete) complete fields can be interpreted in the following fashion. The group $M^{(p)}$ is the minimal set of multiplicative functions over k which suffices to describe all finite regular fields. Somehow the elements ω_i correspond to the unramified algebraic functions—obtained from normalized integrals of third kind—of the classical case. The λ_i correspond to ramified algebraic functions with one prescribed pole. Of course, this analogy is rather superficial.

We now want to interpret the Galois groups G_i of the fields A_i/k in terms of certain factor groups of division algebras over k.

Let D_i be the normal division algebra of degree p^i over k which is given by the following relations. The algebra D_i is to consist of all finite sums $\sum_{j,k=1}^{p^i} a_{j,k}\Omega_i^j\Lambda_i^k$, where $a_{j,k}$ are elements of k and Ω_i, Λ_i satisfy the relations

(i) $\Lambda_i^{p^i}=\lambda$,

(ii) Ω_i is a primitive $(q^{p^i}-1)$st root of unity,

(iii) $\Lambda_i^{-1}\Omega_i\Lambda_i=\Omega_i^q$ (cf. [4]).

Let $\mathfrak{P}_i$ be the two-sided prime ideal of D_i and $\{E_i\}$ the totality of all units in D_i which are congruent to 1 (mod $\mathfrak{P}_i$). Obviously, the set $\{E_i\}$ is a group.

LEMMA 6. *Every element of* $\{E_i\}$ *is a* p^i*th power.*

Proof. Let E_i be an arbitrary element of $\{\mathrm{E}_i\}$. Then $\mathrm{E}_i = 1 + B_i$ where $V_{\mathfrak{p}_i}(B_i) > 0$. Consider the subfield $K(B_i)$ of D_i. It contains E_i. Hence, by the proof of Lemma 1, $\mathrm{E}_i = \mathrm{H}^{p^i}$ where H is a suitable unit which is congruent to 1 modulo the prime ideal of $k(B_i)$. Since $\mathrm{H} \in k(B_i) \subset D_i$, the assertion of the lemma is obvious.

THEOREM 6. *The factor group $D_i^*/D_i^{*p^i}$ can be generated by two elements Λ_i^*, Ω_i^* satisfying the relations*

$$\Lambda_i^{*p^i} = \Omega_i^{*p^i} = 1, \qquad \Lambda_i^{*-1}\Omega_i^*\Lambda_i^* = \Omega_i^{*q}.$$

Proof. We first remark that the group $D_i^{*p^i}$ which is generated by the p^ith powers of all nonzero elements in D_i^* is an invariant subgroup of the multiplicative group D_i^*. The group $D_i^{*p^i}$ contains the unit group $\{\mathrm{E}_i\}$, by Lemma 6. Thus, Λ_i and Ω_i can be considered as representatives of $D_i^*/D_i^{*p^i}$. The defining relations of D_i and the fact that $\{\mathrm{E}_i\}$ is an invariant subgroup of D_i^* imply that every class of $D_i^*/D_i^{*p^i}$ can be represented as $\Lambda_i^{\alpha_i}\Omega_i^{\beta_i}D_i^{*p^i}$ where $0 \leqq \alpha_i, \beta_i < p^i$. Hence $\Lambda_i^* = \Lambda_i \pmod{D_i^{*p^i}}$ and $\Omega_i^* = \Omega_i \pmod{D_i^{*p^i}}$ have the required properties.

We can suppose that the Galois group G_i of A_i/k is given by the relations $\sigma_i^{p^i} = \tau_i^{p^i} = 1$, $\tau_i^{-1}\sigma_i\tau_i = \sigma_i^q$. Namely, $q-1$ is exactly divisible by p^μ.

THEOREM 7. *The Galois group G_i of A_i/k is isomorphic with the factor group $D_i^*/D_i^{*p^i}$ of D_i.*

Proof. The asserted isomorphism is an immediate consequence of Theorem 6 and Lemma 4. We let correspond $\sigma_i \rightarrow \Omega_i^*$, $\tau_i \rightarrow \Lambda_i^*$.

The group $D_i^{*p^i}$ can be considered as the generalization of the norm groups appearing in local class field theory. We can obtain $D_i^{*p^i}$ by the following construction. Consider the norm group $N_iA_i^* = k^{*p^i}$ which belongs to the field A_i/k. Let $\mathfrak{N}_i$ denote the reduced norm of D_i over k. Then $D_i^{*p^i}$ is the subgroup H of maximal index in D_i^* such that $\mathfrak{N}_iH = N_iA_i^*$. This follows immediately if one observes that $\mathfrak{N}_iD_i^* = k^*$, and that Λ_i, Ω_i, $\{\mathrm{E}_i\}$ form a base for all elements of D_i^*. The requirement $\mathfrak{N}_iH = N_iA_i^*$ implies certain congruence conditions for the exponents occurring in the multiplicative representation of the elements in H in terms of the base Λ_i, Ω_i, $\{\mathrm{E}_i\}$. Since $\mathfrak{N}_i$ is an abelian multiplicative function on D_i^*, it follows that H is a normal subgroup of D_i^*.

Now let $D_i^{*\prime}$ be the *maximal* subgroup of D_i^* such that $\mathfrak{N}_iD_i^{*\prime} = N_iA_i^*$. Using the homomorphism principle on groups we get

$$[D_i^* : D_i^{*\prime}] = [D_i^*/D_i^{*p^i} : D_i^{*\prime}/D_i^{*p^i}] = [k^* : N_iA_i^*].$$

The last index is equal to $[G_i : G_i']$, where G_i' denotes the commutator group

of G_i (cf. [1, 8, 10]). Hence $D_i^{*\prime}/D_i^{*p^i}$ is the commutator group of $D_i^*/G_i^{*p^i}$, for $D_i^*/D_i^{*p^i}\cong G_i$ by Theorem 7. Thus, $D_i^{*\prime}/D_i^{*p^i}\cong G_i'$. The latter group is a cyclic group which is generated by the commutator $\tau_i^{-1}\sigma_i\tau_i\sigma_i^{-1}=\sigma_i^{q-1}$. Namely, the structure of A_i/k implies that $A_i=U_{p^i}k(\lambda^{1/p^\mu})(\lambda^{1/p^\mu})$ is a cyclic extension of degree $p^{i-\mu}$ over $U_{p^i}k(\lambda^{1/p^\mu})$. The latter field A_i' obviously belongs to G_i' according to the local class field theory. We have

$$[D_i^*:D_i^{*p^i}] = [\mathfrak{N}_iD_i^*:\mathfrak{N}_iD_i^{*p^i}][X_1:Y_1],$$
$$p^{2i} = [k^*:N_iA_i^*][X_1:Y_1] = p^{i+\mu}p^{i-\mu}, \qquad \text{if } i>\mu.$$

Here X_1 denotes the group of elements in D_i^* whose reduced norms are equal to 1 and similarly, Y_1 the appropriate subgroup of $D_i^{*p^i}$. These statements yield that $D_i^{*\prime}/D^{p^i}$ is represented by a root of unity. For this we only have to take into account the structure of D_i^* and the multiplicative properties of $\mathfrak{N}_i$. It is therefore not so astounding that the Galois groups G_i can be described by factor groups of D_i^*. The algebra D_i contains the $(q^{p^i}-1)$st roots of unity and hence the $p^{\mu+i}$th roots of unity, for the maximal unramified subfield U^{p^i} of D_i has relative degree p^i over k. Consequently U_{p^i} must be given by ω^{1/p^i}, where ω is a primitive $(q-1)$st root of unity in k, for we supposed $q-1\equiv 0\ (p)$.

The field A_i' can be obtained as follows. Consider the group $D_i^*/D_i^{*p^i}$ and reduce it modulo the commutator group $D_i^{*\prime}/D_i^{*p^i}$. Then $\Lambda_i^{*p^\mu}=\lambda^*$ and $\Omega_i^{*p^i}=\omega^*$. Consider λ^* and ω^* as the elements λ, ω in k. Then the solutions x_i, y_i of $x_i^{p^\mu}=\lambda$, $y_i^{p^i}=\omega$ define the field A_i'.

In order to obtain A_i itself, we associate to $D_i^*/D_i^{*p^i}$ the group $\bar{A}_i$ given by the relations $X_i^{p^i}=\lambda$, $Y_i^{p^i}=\omega$. Then $A_i=k(X_i, Y_i)$. This is to a certain extent the generalization of the theory of radical fields to normal regular extensions.

Suppose now that K is an arbitrary regular normal extension of k. Then $K\subseteqq A_i$ for some A_i by Theorem 2. To the field K there corresponds a uniquely determined invariant subgroup $\overline{H}$ of G_i such that $G(K, k)\cong G_i/\overline{H}$. The local class field theory yields that $G(K, k)/G^*(K, k)'\cong k^*/N_KK^*$, where $G(K, k)'$ denotes the commutator group of $G(K, k)$ and N_KK^* stands for the norm class group of K with respect to k. Theorem 7 implies the existence of a unique invariant subgroup H of D_i^* such that $D_i^*/H\cong G(K, k)$.

THEOREM 8. *Let K be an arbitrary normal regular extension of k such that $K\subseteqq A_i$, then $[k^*:\mathfrak{N}_iH]=[k^*:N_KK^*]$.*

Proof. The first isomorphism theorem of group theory yields that there exists an invariant subgroup $\overline{R}$ of G_i such that $G_i/\overline{R}\cong G(K, k)/G(K, k)'$. Hence $\overline{R}\supseteqq G_i'$ by the properties of the commutator group G_i'. Trans-

ferring $\overline{R}$ to D_i^*, we obtain an invariant subgroup R of D_i^* for which $D_i^*/R \cong G(K, k)/G(K, k)' \cong k^*/N_K K^*$; $R \supseteq D_i^{*\prime}$. Next consider the index $[D_i^*:H]$. Applying the multiplicative mapping N_i to D_i^* and H, we get

$$[D_i^*:H] = [\mathfrak{N}_i D_i^*: \mathfrak{N}_i H][D_i^{*\prime}:D_i^{*\prime} \cap H] = [k^*: \mathfrak{N}_i H][D_i^{*\prime}:D_i^{*\prime} \cap H].$$

The second isomorphism theorem of group theory yields

$$D_i^{*\prime}/D_i^{*\prime} \cap H \cong D_i^* \cup H/H.$$

Consequently, $[k^*: \mathfrak{N}_i H] = [D_i^*: H \cup D_i^{*\prime}]$. Namely, the inclusion $D_i^* \supseteq H \cup D_i^{*\prime} \supseteq H$ implies

$$[D_i^{*\prime} \cup H:H] = [D_i^*:H][D_i^*:D_i^{*\prime} \cup H]^{-1}.$$

Thus $[D_i^*:H] = [k^*: \mathfrak{N}_i H][D_i^*:H][D_i^*:D_i^{*\prime} \cup H]^{-1}$. Since D_i^*/R is an abelian group, we have $R \supseteq D_i^{*\prime}$, by the properties of the commutator group $D_i^{*\prime}$. Hence $R \supseteq D_i^{*\prime} \cup H$. But also $D_i^{*\prime} \cup H \subseteq R$ for R/H is the commutator group of D_i^*/H. Consequently, $D_i^{*\prime} \cup H = R$.

Application of N_i to the groups D_i^* and $D_i^* \cup H$ yields

$$\begin{aligned}[D_i^*:D_i^{*\prime} \cup H] &= [\mathfrak{N}_i D_i^*: \mathfrak{N}_i D_i^{*\prime} \cap H][D_i^{*\prime}:D_i^{*\prime} \cup (H \cap D_i^{*\prime})] \\ &= [k^*: \mathfrak{N}_i H].\end{aligned}$$

Finally, $[k^*:N_K K^*] = [k^*: \mathfrak{N}_i H]$ for $D_i^*/R \cong G(K, k)/G(K, k)'$.

Finally we want to indicate briefly that the theory of regular normal extensions K/k cannot be developed with respect to a fixed division algebra D_i of degree p^i over k. It would be natural to require that every Galois group G is a homomorphic map of D_i^*. In analogy to the theory of abelian fields (which can be described by subgroups of k^*) we would require that all units $\{\mathfrak{E}_i\}$ of D_i are contained in the subgroups H of D_i^* belonging to the arbitrarily given field K/k. Namely, the set of units $\mathfrak{E}_i^{(j)} \equiv ((\Lambda_i)^j)$ form an invariant subgroup of the totality of *all* units in D_i. The index of this subgroup turns out to be equal to $q^{p^i j-1}(q^{p^i}-1)$. Thus $p^{\mu p^i} \| q^{p^i j-1}(q^{p^i}-1)$ for D_i contains no units of higher p-order than $p^{\mu p^i}$, as follows from the algebraic theory of splitting fields of D_i. Hence the class groups H which supposedly can be associated to the fields K/k must contain $\mathfrak{E}_i$. Consequently, any such invariant group H of D_i must contain $\{\Omega_i^m\} \cup \{\Lambda_i^s\}$ for suitable chosen integers m, s. Since D_i contains only roots of unity of bounded p-order, the possible exponents are bounded, say $m \leqq m_0$. If $m > m_0$, then $\{\Omega_i^m\} \cup \{\Lambda_i^s\} = \{\Omega_i^{m_0}\} \cup \{\Lambda_i^s\}$. We have a similar fact in local class field theory. There the ramification exponents $e(K)$ of regular fields are bounded by p^μ. The reason being that the ground field k does not contain sufficiently may roots of unity. Such roots of unity always are needed to describe the class groups of ramified abelian extensions. In

other words, in the abelian case the regular extensions are made up by a ramified extension of degree p^μ and an unramified extension of arbitrarily high degree (cf. [10]). A similar situation prevails in the general case. Having fixed an algebra D_i, we only can describe by means of D_i such regular fields K which contain a normal subfield whose ramification degree is not greater than $p^{\mu p^i}$. One also could arrive at this result by purely group theoretical analysis. The structure of $D_i^*/\{\Omega_i^m\} \cup \{\Lambda_i^s\}$ can easily be determined. One observes that not all G_i can be homomorphic maps of some suitable $D_i^*/\{\Omega_i^m\} \cup \{\Lambda_i^s\}$. Hence, by virtue of Theorem 2, not every G can be described as a homomorphic map of D_i^*. These considerations indicate why the totality of all algebras D_i is required for our generalization of local class field theory.

Bibliography

1. Y. Akizuki, *Eine homomorphe Zuordnung der Elemente der galoisschen Gruppe zu den Elementen einer Untergruppe der Normklassengruppe*, Mathematische Annalen, vol. 112 (1936).

2. W. Burnside, *Theory of Groups of Finite Order*.

3. C. Chevalley, *La théorie du symbole de restes normiques*, Journal für die reine und angewandte Mathematik, vol. 169 (1933).

4. H. Hasse, *Über p-adische Schiefkörper und ihre Bedeutung für die Arithmetik hyperkomplexer Zahlsysteme*, Mathematische Annalen, vol. 104 (1930).

5. K. Hensel, (1) *Allgemeine Theorie der Kongruenzklassengruppen und ihrer Invarianten in algebraischen Zahlkörpern*, Journal für die reine und angewandte Mathematik, vol. 147 (1917); (2) *Zahlentheorie*, Berlin and Leipzig.

6. J. Herbrand, *Théorie arithmétique des corps de nombres de degré infini*, II. *Extensions algébriques de degré infini*, Mathematische Annalen, vol. 108 (1933).

7. S. MacLane, *Subfields and automorphisms of p-adic fields*, Annals of Mathematics, (2), vol. 40 (1939).

8. T. Nakayama, *Über die Beziehungen zwischen den Faktorensystemen und der Normklassengruppe eines galoisschen Erweiterungskörpers*, Mathematische Annalen, vol. 112 (1936).

9. A. Ostrowski, *Untersuchungen zur arithmetischen Theorie der Körper*, Mathematische Zeitschrift, vol. 39 (1935).

10. F. K. Schmidt, *Zur Klassenkörpertheorie im Kleinen*, Journal für die reine und angewandte Mathematik, vol. 162 (1930).

11. O. Teichmüller, *Diskret bewertete perfekte Körper mit unvollkommenem Restklassenkörper*, Journal für die reine und angewandte Mathematik, vol. 176 (1936).

12. E. Witt, (1) *Konstruktion von galoisschen Körpern der Charakteristik p mit gegebener Gruppe der Ordnung p^n*, Journal für die reine und angewandte Mathematik, vol. 174 (1936); (2) *Zyklische Körper und Algebren der Charakteristik p vom Grade p^n*, Journal für die reine und angewandte Mathematik, vol. 176 (1936).

University of Chicago,
Chicago, Ill.

REGULARITY PROPERTIES OF CERTAIN FAMILIES OF CHANCE VARIABLES*

BY

J. L. DOOB

Let t vary in any simply ordered set, and let $\{x_t\}$ be a family of chance variables. We shall say that the chance variables have the property $\mathcal{E}$ if whenever $t_1 < \cdots < t_{n+1}$,

$$E[x_{t_1}, \cdots, x_{t_n}; x_{t_{n+1}}] = x_{t_n},\dagger$$

with probability 1. The subscript n here is any positive integer. It will always be supposed that the given chance variables have expectations, so that the conditional expectations involved in the definition of the property $\mathcal{E}$ will always exist.

We shall first make a few introductory remarks on the general family of chance variables with the property $\mathcal{E}$. Then in §1 it will be supposed that t takes on integral values, and the convergence properties of such sequences of chance variables will be discussed in detail. In §3 it will be supposed that t runs through the real numbers. Before discussing this case, it is necessary to investigate the justification for the use of such descriptive terms as *continuity*, *boundedness*, and so on, as applied to the random function x_t. This investigation is made in §2, in the general case, without the hypothesis of the property $\mathcal{E}$, and it is found that the above terms can always be made meaningful, and given their usual meanings, at the cost, however, in some cases, of introducing infinite-valued functions.

In the following, we shall always suppose that the x_t are measurable functions defined on a space Ω, on certain sets of which a measure function is defined. That this can always be done, and how this is to be done, was shown by Kolmogoroff [5, pp. 27–30]. The space Ω, following Kolmogoroff, will be taken to be the space of real-valued functions of t. Integration in terms of probability measure will be denoted by $\int \cdots dP$, and integration will be over all space, unless the domain of integration is otherwise specified. The quali-

* Presented to the Society, December 28, 1939, under the title *On a certain type of family of chance variables*; received by the editors December 11, 1939.

† We shall use the notation $E[y]$ for the expectation of the chance variable y, and $E[y_1, \cdots, y_n; y]$ for the conditional expectation of y for given $y_1, \cdots, y_n$, a function of $y_1, \cdots, y_n$. If the y_j are not finite in number, the notation will be modified accordingly. We shall assume the definitions of Kolmogoroff [5, pp. 41–44] for these conditional expectations. (References in brackets are to the bibliography at the end of the paper.)

fication "with probability 1" will be used interchangeably with "almost everywhere on Ω." Let I be any set of t-values. A P-measurable set Λ (that is, an Ω-set whose probability is defined) will be said to be an x_t-set with $t \in I$, or a set depending only on t for t in I, if it is in the Borel field of Ω-sets determined by the sets of the form

$$\{x_{t_j} < k_j,\ j = 1, \cdots, n\}$$

(n any integer, $k_1, \cdots, k_n$ any n numbers, $t_1, \cdots, t_n$ any n distinct t-values) or if it differs from an Ω-set in this field by an Ω-set of P-measure 0. If I includes all t-values, the Ω-sets in question are all the P-measurable sets.

THEOREM 0.1. *Let $\{x_t\}$ be a family of chance variables with the property $\mathcal{E}$. If I is any t-set, and if $t \leqq s_1, s_2$ for all t in I, then*

$$E[x_t, t \in I; x_{s_1}] = E[x_t, t \in I; x_{s_2}], \tag{0.1.1}$$

with probability 1. *If I has a last element s, and if $\sigma > s$,*

$$E[x_t, t \in I; x_\sigma] = x_s, \tag{0.1.2}$$

with probability 1.

This theorem extends the defining property $\mathcal{E}$, which states the conclusion of Theorem 0.1 for I a finite set. By definition of conditional expectations, $E[x_t, t \in I; x_{s_j}]$, $j=1, 2$, is a function measurable with respect to the P-measurable Ω-sets depending only on x_t for t in I, and such that, if Λ is an Ω-set of this type,

$$\int_\Lambda E[x_t, t \in I; x_{s_j}]dP = \int_\Lambda x_{s_j}dP, \qquad j = 1, 2. \tag{0.1.3}$$

Suppose that the P-measurable set M depends only on t for $t = t_1, \cdots, t_n$, with $t_j \in I$, $j \leqq n$, $t_1 < \cdots < t_n$. Then, using the fact that the chance variables $x_{t_1}, \cdots, x_{t_n}, x_{s_j}$ have the property $\mathcal{E}$,

$$\int_M x_{s_j}dP = \int_M E[x_{t_1}, \cdots, x_{t_n}; x_{s_j}]dP = \int_M x_{t_n}dP, \qquad j = 1, 2. \tag{0.1.4}$$

Thus

$$\int_M x_{s_1}dP = \int_M x_{s_2}dP, \tag{0.1.5}$$

and therefore, using (0.1.3),

$$\int_M E[x_t, t \in I; x_{s_1}]dP = \int_M E[x_t, t \in I; x_{s_2}]dP. \tag{0.1.6}$$

Since any P-measurable set which depends only on t for $t \in I$ can be approximated arbitrarily closely by a set M, (0.1.6) must hold with the general Λ in place of M. Since the integrands are measurable with respect to the field of sets Λ, the integrands are necessarily equal, with probability 1, as was to be proved. If I has a last element s, $E[x_t, t \in I; x_s] = x_s$, with probability 1, from the definition of conditional expectations, so in that case both terms of (0.1.1) are equal to $E[x_t, t \in I; x_s] = x_s$, with probability 1. The last statement of the theorem is merely a rewording of this fact.

THEOREM 0.2. *Let x, y be chance variables whose expectations exist, and suppose that $E[x; y] = x$, with probability* 1. *Then $E[x] = E[y]$, and $E|x| \geqq E|y|$.*

If $E[x; y] = x$, with probability 1, the expectations of the two chance variables $E[x; y]$, x are equal: $E[y] = E[x]$. Moreover $|x| \geqq E[x; |y|]$ with probability 1, and, taking expectations of the two sides of this inequality, we obtain $E|x| \geqq E|y|$.

COROLLARY. *If $\{x_t\}$ is a family of chance variables with the property $\mathcal{E}$, Ex_t is independent of t, and $E|x_t|$ is a non-decreasing function of t.*

THEOREM 0.3. *Let $\{x_t\}$ be a family of chance variables with the property $\mathcal{E}$. Then for each t_0, the x_t with $t \leqq t_0$ are uniformly integrable.**

We show first that

$$\lim_{k \to \infty} P\{|x_t| \geqq k\} = 0, \tag{0.3.1}$$

uniformly for $t \leqq t_0$. This fact is implied by the following inequality (in which $t \leqq t_0$):

$$\begin{aligned} kP\{|x_t| \geqq k\} &\leqq \int_{\{x_t \geqq k\}} x_t dP - \int_{\{x_t \leqq -k\}} x_t dP \\ &= \int_{\{x_t \geqq k\}} x_{t_0} dP - \int_{\{x_t \leqq -k\}} x_{t_0} dP \leqq \int |x_{t_0}|\, dP. \end{aligned} \tag{0.3.2}$$

We are using here the fact that if $t \leqq t_0$, $E[x_t; x_{t_0}] = x_t$ with probability 1. Now if $t \leqq t_0$,

$$\begin{aligned} \int_{\{|x_t| \geqq k\}} |x_t|\, dP &= \int_{\{x_t \geqq k\}} x_t dP - \int_{\{x_t \leqq -k\}} x_t dP \\ &= \int_{\{x_t \geqq k\}} x_{t_0} dP - \int_{\{x_t \leqq -k\}} x_{t_0} dP \leqq \int_{\{|x_t| \geqq k\}} |x_{t_0}|\, dP, \end{aligned} \tag{0.3.3}$$

* This means that $\lim_{k \to \infty} \int_{\{|x_t| \geqq k\}} |x_t| dP = 0$ uniformly in t for $t \leqq t_0$.

and, as $k \to \infty$, the right side goes to 0 uniformly in t (from (0.3.1)), so the left side goes to 0 uniformly in t, as was to be proved.

1. **Sequences of chance variables with the property $\mathcal{E}$.** We first prove the following theorem.

THEOREM 1.1. *Let $\cdots, x_{-1}, x_0, \cdots, x$ be chance variables with the property $\mathcal{E}$.* Let m, n be integers with $m \leqq n$, and define the sets Λ, M by*

$$\Lambda = \left\{ \underset{m \leqq j \leqq n}{\text{L.U.B.}}\ x_j \geqq k \right\}, \qquad \mathrm{M} = \left\{ \underset{m \leqq j \leqq n}{\text{G.L.B.}}\ x_j \leqq k \right\}.$$

Then if N *is any P-measurable $(\cdots, x_{m-1}, x_m)$-set,*

$$(1.1.1) \qquad \int_{\Lambda \cdot \mathrm{N}} x dP \geqq kP\{\Lambda \cdot \mathrm{N}\}, \qquad \int_{\mathrm{M} \cdot \mathrm{N}} x dP \leqq kP\{\mathrm{M} \cdot \mathrm{N}\}. \dagger$$

To prove (1.1.1), define the sets $\Lambda_m, \cdots$ as follows:

$$\begin{aligned} \Lambda_m &= \{x_m \geqq k\}, \\ \Lambda_{m+1} &= \{x_m < k,\ x_{m+1} \geqq k\}, \cdots, \\ \Lambda_n &= \{x_m < k,\ x_{m+1} < k, \cdots, x_{n-1} < k,\ x_n \geqq k\}. \end{aligned}$$

Then the Λ_j are disjunct with sum Λ, and Λ_j depends only on $x_m, \cdots, x_j$. Since $E[\cdots, x_j; x] = x_j$, with probability 1 (Theorem 0.1),

$$(1.1.2) \qquad \begin{aligned} \int_{\Lambda \cdot \mathrm{N}} x dP &= \sum_m^n \int_{\Lambda_j \cdot \mathrm{N}} x dP = \sum_m^n \int_{\Lambda_j \cdot \mathrm{N}} x_j dP \\ &\geqq k \sum_m^n P(\Lambda_j \cdot \mathrm{N}) = kP(\Lambda \cdot \mathrm{N}), \end{aligned}$$

giving the first inequality of (1.1.1), and the second can be proved in the same way, or by changing the x_j, x to their negatives, and applying the result already obtained.

THEOREM 1.2. *Let $\cdots, x_{-1}, x_0$ be a sequence of chance variables with the property $\mathcal{E}$. Then $\lim_{n \to -\infty} x_n = x$ exists with probability 1, and the chance variables $x, \cdots, x_{-1}, x_0$ have the property $\mathcal{E}$. The chance variables $\{x_j\}$ are uniformly integrable, and $E|x_0| \geqq E|x_{-1}| \geqq \cdots \geqq E|x|$; $E|x_n| \to E|x|$.*

Unless $\liminf_{n \to -\infty} x_n = \limsup_{n \to -\infty} x_n$ with probability 1, there are numbers k_1, k_2 with $k_2 > k_1$, such that

* The x_j may be finite or infinite in number, and if infinite in number may be finite in one direction. The essential fact is that there is a last one x.

† The inequalities (1.1.1) are implicit in the work of Ville [9, pp. 100–101] who discussed sequences of non-negative chance variables with the property $\mathcal{E}$. The method of proof we use was used by Lévy [8, p. 129], in a related discussion.

$$\liminf_{n\to-\infty} x_n < k_1 < k_2 < \limsup_{n\to-\infty} x_n, \tag{1.2.1}$$

with positive probability η. Let the $(\cdots, x_0)$-set on which (1.2.1) is true be Λ. We can choose $-n_1$ so large that L.U.B.$_{n_1\leqq j}\{x_j\}\geqq k_2$ on an $(\cdots, x_0)$-set Λ_1 with $P(\Lambda\cdot\Lambda_1)>\eta(1-2^{-1})$. We can choose $n_2<n_1$ such that G.L.B.$_{n_2\leqq j\leqq n_1}\{x_j\}\leqq k_1$ on an $(\cdots, x_0)$-set Λ_2 with $P(\Lambda\cdot\Lambda_2)>\eta(1-2^{-2})$. In general, if $n_1, \cdots, n_{r-1}$ have already been chosen, and if r is odd, we can choose $n_r<n_{r-1}$ such that L.U.B.$_{n_r\leqq j\leqq n_{r-1}}\{x_j\}\geqq k_2$ on an $(\cdots, x_0)$-set Λ_r with $P(\Lambda\cdot\Lambda_r)>\eta(1-2^{-r})$; or if r is even, we can choose $n_r<n_{r-1}$ such that G.L.B.$_{n_r\leqq j\leqq n_{r-1}}\{x_j\}\leqq k_1$ on an $(\cdots, x_0)$-set Λ_r with $P(\Lambda\cdot\Lambda_r)>\eta(1-2^{-r})$. Then if $\Lambda_n'=\prod_n^\infty\Lambda_j$,

$$P(\Lambda_n') > \eta - \eta 2^{-n} - \eta 2^{-n-1} - \cdots = \eta(1-2^{-n+1}). \tag{1.2.2}$$

According to Theorem 1.1, if r is odd,

$$\int_{\Lambda_r'} x_0 dP \geqq k_2 P(\Lambda_r') \tag{1.2.3}$$

and if r is even

$$\int_{\Lambda_r'} x_0 dP \leqq k_1 P(\Lambda_r'). \tag{1.2.4}$$

As $r\to\infty$, Λ_r' increases to a set Λ' of measure not less than η; (1.2.3) and (1.2.4) become

$$k_2 P(\Lambda') \leqq \int_{\Lambda'} x_0 dP \leqq k_1 P(\Lambda'), \tag{1.2.5}$$

which is impossible, because $k_1<k_2$. We have thus proved that $\lim_{n\to-\infty} x_n = x$ exists, with probability 1. According to Theorem 0.3, the x_j are uniformly integrable. Then their limit x is integrable, and term by term integration of x_j or $|x_j|$ is legitimate. This means that $E|x_n|\to E|x|$. The fact that $E|x_0|\geqq E|x_{-1}|\geqq\cdots$ follows from the corollary to Theorem 0.2. To show that the chance variables $x, \cdots, x_{-1}, x_0$ have the property $\mathcal{E}$, it is only necessary to show that if $m\leqq n\leqq r\leqq 0$, $E[x, x_m, \cdots, x_n; x_r]=x_n$, with probability 1, and that $E[x; x_r]=x$, with probability 1. To prove the first, we must prove that if Λ is a set depending only on $x, x_m, \cdots, x_n$, then

$$\int_\Lambda x_n dP = \int_\Lambda x_r dP. \tag{1.2.6}$$

Now Λ evidently depends only on $\cdots, x_{n-1}, x_n$; so (1.2.6) follows from the fact that $E[\cdots, x_n; x_r]=x_n$, with probability 1 (Theorem 0.1). If Λ is a set

depending only on x, (1.2.6) is true for all $n \leqq r$, and becomes, if $n \to -\infty$,

$$\int_A x dP = \int_A x_r dP,$$

that is, $E[x; x_r] = x$, with probability 1.

THEOREM 1.3. *Let $x_1, x_2, \cdots$ be a sequence of chance variables with the property $\mathcal{E}$. Then $E|x_1| \leqq E|x_2| \leqq \cdots$. If $\lim_{n\to\infty} E|x_n| = l < \infty$, then $\lim_{n\to\infty} x_n = x$ exists, with probability 1, and $E|x| \leqq l$. If the x_j are uniformly integrable, $\lim_{n\to\infty} x_n = x$ exists, with probability 1, and the chance variables $x_1, x_2, \cdots, x$ have the property $\mathcal{E}$.*

Ville has studied sequences of non-negative chance variables with the property $\mathcal{E}$. Since, by the corollary to Theorem 0.2, Ville's hypotheses imply that

$$Ex_1 = Ex_2 = \cdots = E|x_1| = E|x_2| = \cdots,$$

the hypotheses of the first part of Theorem 1.3 are satisfied, in Ville's case. Ville proved* that in his case $\text{L.U.B.}_{j\geqq 1} |x_j| < \infty$, with probability 1 (implied by our conclusion that $\lim_{n\to\infty} x_n$ exists with probability 1, and that the limit is integrable), and applied this fact to the study of certain games of chance.

Proof of Theorem 1.3. According to the corollary to Theorem 0.2, $E|x_n|$ is monotone non-decreasing. Suppose that $\lim_{n\to\infty} E|x_n| = l < \infty$. Unless $\liminf_{n\to\infty} x_n = \limsup_{n\to\infty} x_n$, with probability 1, there are numbers k_1, k_2, with $k_1 < k_2$, such that

$$\liminf_{n\to\infty} x_n < k_1 < k_2 < \limsup_{n\to\infty} x_n \tag{1.3.1}$$

with positive probability η. Let the $(x_1, x_2, \cdots)$-set on which this is true be Λ. We can choose n_1 so large that $\text{L.U.B.}_{j\leqq n_1} \{x_j\} \geqq k_2$ on an $(x_1, x_2, \cdots)$-set Λ_1 with $P(\Lambda \cdot \Lambda_1) > \eta(1-3^{-1})$. We can then choose $n_2 > n_1$ so large that $\text{G.L.B.}_{n_1\leqq j\leqq n_2} \{x_j\} \leqq k_1$ on an $(x_1, x_2, \cdots)$-set Λ_2 with $P(\Lambda \cdot \Lambda_2) > \eta(1-3^{-2})$. In general if $n_1, \cdots, n_{r-1}$ have already been chosen, and if r is odd, we can find $n_r > n_{r-1}$ so large that $\text{L.U.B.}_{n_{r-1}\leqq j\leqq n_r} \{x_j\} \geqq k_2$ on an $(x_1, x_2, \cdots)$-set Λ_r with $P(\Lambda \cdot \Lambda_r) > \eta(1-3^{-r})$; or if r is even, we can find $n_r > n_{r-1}$ so large that $\text{G.L.B.}_{n_{r-1}\leqq j\leqq n_r} \{x_j\} \leqq k_1$ on an $(x_1, x_2, \cdots)$-set Λ_r with $P(\Lambda \cdot \Lambda_r) > \eta(1-3^{-r})$. Then if $\Lambda_n' = \prod_1^n \Lambda_j$,

$$P(\Lambda_n') > \eta - \eta/3 - \eta/3^2 - \cdots = \eta/2. \tag{1.3.2}$$

According to Theorem 1.1, if $m \geqq n_{2r}$,

* [9, chaps. 4, 5.]

$$(1.3.3)\qquad \int_{\Lambda'_{2r}} x_m dP \leqq k_1 P(\Lambda'_{2r}), \qquad \int_{\Lambda'_{2r-1}} x_m dP \geqq k_2 P(\Lambda'_{2r-1}).$$

Then if $M_r = \Lambda'_{2r-1} - \Lambda_{2r}$

$$(1.3.4)\qquad \int_{M_r} x_m dP \geqq k_2 P(\Lambda'_{2r-1}) - k_1 P(\Lambda'_{2r}) = (k_2 - k_1) P(\Lambda'_{2r-1}) + k_1 P(M_r).$$

Now the M_r are disjunct, so if m is sufficiently large (depending on q),

$$(1.3.5)\qquad \begin{aligned} \sum_{r=1}^{q} \int_{M_r} x_m &\geqq (k_2 - k_1) \sum_{r=1}^{q} P(\Lambda'_{2r-1}) + k_1 \sum_{r=1}^{q} P(M_r) \\ &\geqq \frac{q(k_2 - k_1)}{2} \eta - |k_1|. \end{aligned}$$

But this is impossible, since then

$$(1.3.6)\qquad l \geqq \int |x_m| \, dP \geqq \sum_{r=1}^{q} \int_{M_r} x_m dP \geqq \frac{q\eta}{2} (k_2 - k_1) - |k_1|,$$

that is,

$$(1.3.7)\qquad l \geqq (1/2) q\eta (k_2 - k_1) - |k_1|,$$

for all q. We have thus proved that $\lim_{n\to\infty} x_n = x$ exists with probability 1. By Fatou's lemma, $E|x| \leqq l$, and the fact that $E|x| < \infty$ implies that x is finite, with probability 1. If the x_j are supposed uniformly integrable, it follows that the sequence $\{E|x_j|\}$ is uniformly bounded. Then $\lim_{n\to\infty} x_n = x$ exists with probability 1. Because of the uniform integrability, term by term integration of the sequence $\{x_n\}$ is allowed on any P-measurable set, and we can readily verify that the sequence of chance variables $x_1, x_2, \cdots, x$ has the property $\mathcal{E}$. In fact, all that need be proved is that $E[x_{\alpha_1}, \cdots, x_{\alpha_n}; x] = x_{\alpha_n}$ with probability 1, if $\alpha_1 < \cdots < \alpha_n$. This will follow if it can be shown that whenever Λ is an $(x_1, x_2, \cdots)$-set depending only on $x_{\alpha_1}, \cdots, x_{\alpha_n}$, then

$$(1.3.8)\qquad \int_{\Lambda} x dP = \int_{\Lambda} x_{\alpha_n} dP.$$

Since the chance variables $x_1, x_2, \cdots$ have the property $\mathcal{E}$, (1.3.8) is true if x is replaced by x_m with $m > \alpha_n$. Then if $m \to \infty$, we obtain (1.3.8).

COROLLARY. *Let $y_1, y_2, \cdots$ be a sequence of chance variables whose expectations exist, and suppose that $E[y_1, \cdots, y_n; y_{n+1}] = 0$, $n = 1, 2, \cdots$, with probability 1. Then $E|\sum_1^n y_j|$ is monotone non-decreasing with n, and if $\lim_{n\to\infty} E|\sum_1^n y_j| < \infty$, the series $\sum_1^\infty y_j$ is convergent, with probability 1.*

For if $x_n=\sum_1^n y_j$, $n=1, 2, \cdots$, the x_n have the property $\mathcal{E}$, and Theorem 1.3 is applicable, giving the result of the corollary.

Lévy [8, pp. 247–248] has discussed series of chance variables $\sum_1^\infty y_j$, with $E[y_1, \cdots, y_n; y_{n+1}]=0$, $n=1, 2, \cdots$, with probability 1, relating the convergence of the given series to that of their conditional dispersions.

An important particular case of this corollary is that in which the y_j are mutually independent. The hypothesis on the conditional expectations becomes in this case the hypothesis that $Ey_1=Ey_2=\cdots=0$, and the corollary has been deduced in this case.* In this special case it has also been shown that the partial sums $\{x_j\}$ are necessarily uniformly bounded by a chance variable whose expectation exists. It follows that in this case the x_j are always uniformly integrable.

The following example shows that the hypotheses of the first part of Theorem 1.3 do not imply that the sequence $x_1, x_2, \cdots, x$ has the property $\mathcal{E}$. Let $x_0=1$ with probability 1. Let $x_1=1/2$ with probability $(1-4^{-1})$ and $x_1=5/2$ with probability $1/4$. In general if $x_1, \cdots, x_r$ have already been determined, we determine x_{r+1} as follows. The values of x_{r+1} depend only on those of x_r, and if $x_r=a$, we set $x_{r+1}=2^{-r-1}$ with probability $(1-4^{-r-1})$ and $x_{r+1}=\xi$ with probability 4^{-r-1}, choosing ξ to make the expectation of x_{r+1} (with $x_r=a$) equal to 0:

$$\frac{\xi}{4^{r+1}}+\frac{1}{2^{r+1}}\left(1-\frac{1}{4^{r+1}}\right)=a.$$

Then it is easily verified that $x_n=2^{-n}$ with probability $(1-4^{-n})$, if $n\geqq 1$, so that $x_n\to x=0$, with probability 1. The chance variables are all positive and the hypotheses of the first part of Theorem 1.3 are satisfied. The sequence $x_1, x_2, \cdots, x$ certainly does not have the property $\mathcal{E}$.

The following theorem is essentially the converse of the second part of Theorem 1.3.

THEOREM 1.4. *Let $x_1, x_2, \cdots, x'$ be chance variables with the property $\mathcal{E}$. Then*

(a) $\lim_{n\to\infty} x_n=E[x_1, x_2, \cdots; x']$ *with probability* 1;

(b) *the chance variables $x_1, x_2, \cdots, x, x'$ have the property $\mathcal{E}$;*

(c) *if x' is a function of $x_1, x_2, \cdots$, $\lim_{n\to\infty} x_n=x'$, with probability* 1.

Lévy [8, p. 129] proved part (c) of this theorem, under the hypothesis that x only takes on the values 1 and 0. His proof can be extended to cover Theorem 1.4. He puts the result in the following form (we are dropping his restriction on the values taken on by x'); let $y_1, y_2, \cdots$ be a sequence of

* Doob [2].

chance variables and let x' be a chance variable dependent upon them. Then if $x_n = E[y_1, \cdots, y_n; x']$, he proves that $x_n \to x'$ with probability 1. Since $x_1, x_2, \cdots, x'$ have the property $\mathcal{E}$, this is exactly our Theorem 1.4(c).

Proof of (a), (b). According to Theorem 0.3, the x_i are uniformly integrable. Then Theorem 1.3 states that $\lim_{n\to\infty} x_n = x$ exists with probability 1, and that the chance variables $x_1, \cdots, x$ have the property $\mathcal{E}$. If we show that the chance variables $x_1, \cdots, x, x'$ have the property $\mathcal{E}$, it will follow (because of Theorem 0.1) that $x = E[x_1, \cdots; x']$ with probability 1. To show that the chance variables $x_1, \cdots, x, x'$ have the property $\mathcal{E}$ it is sufficient to show that if $\alpha_1 < \cdots < \alpha_n$, then $E[x_{\alpha_1} \cdots, x_{\alpha_n}, x; x'] = x$ with probability 1, that is, that if Λ is any set depending on $x_{\alpha_1}, \cdots, x_{\alpha_n}, x$,

$$\int_\Lambda x' dP = \int_\Lambda x dP. \tag{1.4.1}$$

Now Λ is a set depending on $x_1, x_2, \cdots$ (because $x_n \to x$ with probability 1). Equation (1.4.1) is certainly true for Λ depending only on a finite number of the x_i (because if x_ν is the last of these x_i, both integrals are equal to the integral of x_ν over Λ: $x_1, \cdots, x'$ and $x_1, \cdots, x$ have the property $\mathcal{E}$). Then (1.4.1) is true in the general case because the general Λ can be approximated arbitrarily closely by the special ones.

Proof of (c). If x' is a function of $x_1, x_2, \cdots$, $E[x_1, x_2, \cdots; x'] = x'$ with probability 1, so that $x = x'$ with probability 1.

THEOREM 1.5. *Let $x_1, x_2, \cdots, x$ be chance variables with the property $\mathcal{E}$. Suppose that $E|x \log |x|| < \infty$. There is then a number K depending only on $E|x \log |x||$, and such that $E[\text{L.U.B.}_{n\geqq 1} |x_n|] < K$.*

Let $\xi = \text{L.U.B.}_{n\geqq 1} \{x_n\}$, and let $\Lambda = \{\xi \geqq k\}$, $k > 0$. According to Theorem 1.1,

$$\int_\Lambda x dP \geqq kP(\Lambda). \tag{1.5.1}$$

Now

$$\int_{\Lambda\cdot\{|x|<k/2\}} |x| \, dP \leqq \frac{k}{2} P(\Lambda), \tag{1.5.2}$$

so

$$\int_{\{|x|\geqq k/2\}} |x| \, dP \geqq \int_{\Lambda\cdot\{|x|\geqq k/2\}} x dP \geqq \frac{k}{2} P(\Lambda) = \frac{k}{2} P\{\xi \geqq k\}. \tag{1.5.3}$$

This inequality, together with the fact that $E|x \log |x|| < \infty$, has been

used in an analogous discussion (Wiener [10, pp. 17–18]) to show that $E\xi \leqq K_1$, where K_1 only depends on $E|x \log |x||$. Applying this result to $-x_1, -x_2, \cdots, -x$, completes the proof of the theorem.

We have seen above in the discussion of the corollary to Theorem 1.3 that if the x_j are the partial sums of a series of mutually independent chance variables having zero expectations, the hypothesis that $E|x \log |x|| < \infty$ is unnecessary in Theorem 1.5.

2. **One-parameter families of chance variables.** In this section we shall consider a one-parameter family of chance variables $\{x_t\}$, $-\infty < t < \infty$, or $0 \leqq t < \infty$. The probability relations are those derived from certain elementary probabilities: if $t_1, \cdots, t_n$ are distinct t-values, and if $k_1, \cdots, k_n$ are any real numbers, $P\{x_{t_j} < k_j, j \leqq n\}$ is given. The mathematical setup appropriate to the study of the probability relations of the x_t is the following. Let Ω^* be the space of all real-valued functions of t. A probability measure is determined on Ω^* by the measures of certain elementary sets, which we shall call neighborhoods: a set Λ of functions $x(t)$ satisfying

$$a_j < x(t_j) < b_j, \qquad j = 1, \cdots, n, \quad -\infty \leqq a_j < b_j \leqq +\infty,$$

will be called a neighborhood, and we define a measure of Λ, $P^*(\Lambda)$, setting

$$P^*(\Lambda) = P\{a_j < x_{t_j} < b_j, j \leqq n\}.\dagger$$

If $x_s(\omega)$ is the function of $\omega \in \Omega^*$ which for $\omega = \omega_0$: $x_0(t)$ assumes the value $x_0(s)$ (so that we can write, somewhat loosely, $x_s(\omega) = x(s)$), $x_s(\omega)$ is a P^*-measurable function (s fixed), and the probability relations of the $\{x_t\}$ become measure relations of the corresponding $\{x_t(\omega)\}$. The outer measure $\overline{P}^*(\Lambda)$ of an Ω^*-set Λ is defined as the lower limit of $P^*(\Lambda')$ where $\Lambda' \supset \Lambda$ and Λ' is P^*-measurable.‡ The inner measure of an Ω^*-set Λ is defined as $1 - \overline{P}^*(\Lambda_1)$, where Λ_1 is the complement of Λ. It is usually convenient to use a subspace Ω of Ω^* (with $\overline{P}^*(\Omega) = 1$) instead of Ω^* itself. The P^*-measure on Ω^* determines P-measure on Ω, and

$$P^*\{x(t_j) < k_j, j = 1, \cdots, n\} = P\{x(t_j) < k_j, j = 1, \cdots, n\}.$$

We are using $x(t_j)$ to denote the P-measurable function $x_{t_j}(\omega)$, which in turn represents the original chance variable x_{t_j}, and we shall continue to use this notation where there is no danger of confusion. The symbol $x(t)$ thus may represent either a point ω of Ω^*, that is, a function of t, or, if we are fixing t, a function of ω: $x_t(\omega)$.

In discussions of continuity and related questions, it is desirable to con-

† Cf. Doob [1, pp. 107–110].

‡ We can even suppose that Λ' is restricted to be a denumerable sum of neighborhoods.

sider L.U.B.$_{t \in I} x_t$, G.L.B.$_{t \in I} x_t$, where I is a t-interval. It was pointed out by Kolmogoroff [5, p. 26] that the probability that L.U.B.$_{t \in I} x_t \leqq k$ or that G.L.B.$_{t \in I} x_t \leqq k$ cannot be derived from the given probabilities. Since we are giving a complete treatment of the problem here, we shall prove a somewhat stronger result, pointed out to the writer by Dr. P. R. Halmos.

THEOREM 2.1. *Let Ω^* and P^*-measure be defined as explained above. Then the ω-set $\{\text{L.U.B.}_{t \in I} x(t) \leqq k\}$ has inner measure 0, for any nondenumerable set I and real number k.*

This theorem shows that the ω-set $\{\text{L.U.B.}_{t \in I} x(t) \leqq k\}$, if it is P^*-measurable, must have measure 0, irrespective of the P^*-measure. If the complement of the ω-set in question is Λ, it will be sufficient to show that $\overline{P}^*(\Lambda) = 1$. The Ω^*-set Λ is the set of all $x(t)$ with $x(t_0) > k$ for some t_0 in I. Let $\Gamma = \Omega^*$ be any denumerable sum of neighborhoods, with $\Gamma \supset \Lambda$. It will be sufficient to show that $P^*(\Gamma) = 1$, and in fact we shall show that $\Gamma = \Omega^*$. Let $t_1, t_2, \cdots$ be the t-values used in defining the Γ_j. Let $x_0(t)$ be any element of Ω^*. There is surely a function $x_1(t)$, with $x_1(t_j) = x_0(t_j)$ for all j, and such that for some s in I, $x_1(s) > k$. Then $x_1(t) \in \Lambda \subset \Gamma$, so $x_0(t) \in \Gamma$ also, because $x_1(t_j) = x_0(t_j)$ $(j \geqq 1)$. Thus $\Gamma = \Omega^*$, as was to be proved.

Two ways have been suggested, to define $P\{\text{L.U.B.}_{t \in I} x_t \leqq k\}$ and $P\{\text{G.L.B.}_{t \in I} x_t \leqq k\}$. Khintchine [4, pp. 68–69] considers a sequence $\{t_j\}$ of t-values in I. The probability $P\{\text{L.U.B.}_{j \geqq 1} x_{t_j} \leqq k\}$ is already determined, and Khintchine defines $P\{\text{L.U.B.}_{t \in I} x_t \leqq k\}$ as the greatest lower bound of these probability numbers, for all denumerable sets $\{t_j\}$ in I; $P\{\text{G.L.B.}_{t \in I} x_t \leqq k\}$ can be defined similarly. With this definition, the "L.U.B." in "L.U.B.$_{t \in I} x_t$" is not taken literally, necessarily not, since x_t is not a single function of t and since the problem has not been reduced to one of studying functions of t. The second way is to find a subspace Ω of Ω^*, of outer measure 1, such that there is a denumerable set $\{s_j\}$ for which, if $x(t) \in \Omega$, and if I is an open interval,

$$(2.1.1) \qquad \underset{t \in I}{\text{L.U.B.}}\, x(t) = \underset{s_j \in I}{\text{L.U.B.}}\, x(s_j), \qquad \underset{t \in I}{\text{G.L.B.}}\, x(t) = \underset{s_j \in I}{\text{G.L.B.}}\, x(s_j).$$

If there is such an Ω, the $x(t)$-set $\{\text{L.U.B.}_{t \in I} x(t) \leqq k\}$ becomes a P-measurable set:

$$(2.1.2) \qquad P\left\{\underset{t \in I}{\text{L.U.B.}}\, x(t) \leqq k\right\} = P\left\{\underset{s_j \in I}{\text{L.U.B.}}\, x(s_j) \leqq k\right\},$$

and similarly for the G.L.B. Evidently the two methods of definition give the same numerical values to $P\{\text{L.U.B.}_{t \in I} x_t \leqq k\}$. It must be stressed that in (2.1.2) the set $\{\text{L.U.B.}_{t \in I} x(t) \leqq k\}$ is a subset of Ω, whose P-measure is equal

to that of the set on the right since the sets are identical. The P-measure of the set on the right is equal to the P^*-measure of the Ω^*-set $\{\text{L.U.B.}_{s_j \in I}\, x(s_j) \leqq k\}$, by the definition of P-measure, and may be any number between 0 and 1, depending on k and the particular characteristics of the given P^*-measure. On the other hand, Theorem 2.1 shows that the Ω^*-set $\{\text{L.U.B.}_{t \in I}\, x(t) \leqq k\}$ necessarily has inner measure 0.

A process with a space Ω as just described is called quasi-separable. Let I be any open t-interval (of length $|I|$) containing the point s. Define $x_*(s)$, $x^*(s)$ by

$$(2.1.3) \qquad x_*(s) = \lim_{|I|\to 0} \text{G.L.B.}_{s_j \in I}\, x(s_j), \qquad x^*(s) = \lim_{|I|\to 0} \text{L.U.B.}_{s_j \in I}\, x(s_j).$$

Then the condition that Ω be the space of a quasi-separable process is equivalent to the condition that if $x(t) \in \Omega$, and if $\{s_j\}$ is an everywhere dense denumerable t-set,

$$(2.1.4) \qquad x_*(t) \leqq x(t) \leqq x^*(t)$$

for all t.† It has been shown (Doob [1]) that many given probability measures on Ω^* are such that there is a quasi-separable process. It will be shown below that every probability measure has this property, if infinite-valued functions are allowed. Applications will then be made in §3 to P^*-measures in which the chance variables $\{x_t\}$ have the property $\mathcal{E}$.

In order to treat the problems below it will sometimes be necessary to allow our functions $x(t)$ to take on the values $\pm\infty$. Kolmogoroff's proof [5, pp. 27–30] that the elementary given probabilities determine a probability measure needs no change to cover this case. All distributions will be finite: for each t, $|x(t)| < \infty$ with probability 1. It is easily shown that the set of everywhere finite-valued functions $x(t)$ has outer measure 1, and inner measure 0.

Let I be an open t-interval, and let $\{t_j\}$ be a sequence in I. Then $\text{L.U.B.}_{j\geqq 1}\, x(t_j)$ is a chance variable, that is, a P^*-measurable function on Ω^*. We shall show below that there is a sequence $\{\bar{t}_j\}$ in I such that if $\{t_j\}$ is any sequence in I,

$$\xi = \text{L.U.B.}_{j\geqq 1}\, x(\bar{t}_j) \geqq \text{L.U.B.}_{j\geqq 1}\, x(t_j)$$

with probability 1. The (possibly infinite-valued) chance variable ξ will be called the generalized upper bound of $x(t)$ in I, and will be denoted by $U_I[x(t)]$. The generalized lower bound $L_I[x(t)]$ is defined in a similar fashion. To demonstrate the existence of the $\{\bar{t}_j\}$, we note that the values of

† Cf. Doob [1, pp. 110–111].

$E[\arctan \text{L.U.B.}_{j\geqq 1}\, x(t_j)]$† for $\{t_j\}$ a sequence in I, have a least upper bound M. There are sequences $\{t_j^{(k)}\}$, $k=1, 2, \cdots$, such that

$$E\left[\arctan \underset{j\geqq 1}{\text{L.U.B.}}\, x(t_j^{(k)})\right] \to \text{M}.$$

Let $\bar{t}_1, \bar{t}_2, \cdots$ be the $t_j^{(k)}$, $j, k \geqq 1$, arranged in some order. Then evidently

$$E\left[\arctan \underset{j\geqq 1}{\text{L.U.B.}}\, x(\bar{t}_j)\right] = \text{M}.$$

Now let $\{t_j\}$ be any sequence in I. We shall show that it is impossible that $\text{L.U.B.}_{j\geqq 1}\, x(t_j) > \text{L.U.B.}_{j\geqq 1}\, x(\bar{t}_j)$ with positive probability, that is, on an Ω^*-set of positive P^*-measure. If this inequality were possible with positive probability, the inequality would still be true if the sequence $\{t_j\}$ were augmented to include the $\bar{t}_j$. Then, supposing this has been done,

$$\arctan \underset{j\geqq 1}{\text{L.U.B.}}\, x(t_j) \geqq \arctan \underset{j\geqq 1}{\text{L.U.B.}}\, x(\bar{t}_j)$$

for all ω, and there is actual inequality with positive probability. This implies that

$$E\left[\arctan \underset{j\geqq 1}{\text{L.U.B.}}\, x(t_j)\right] > E\left[\arctan \underset{j\geqq 1}{\text{L.U.B.}}\, x(\bar{t}_j)\right] = \text{M},$$

contradicting the definition of M. Thus the $\bar{t}_j$ have the required property. Evidently $U_I[x(t)]$ is not a uniquely determined function, but is determined up to an Ω^*-set of P^*-measure 0, in the sense that two generalized upper bounds are equal almost everywhere on Ω^*.

THEOREM 2.2. *There is a sequence* $\{s_j\}$, *independent of the open interval* I, *such that, for each* I,

$$(2.2.1) \qquad \underset{s_j \in I}{\text{L.U.B.}}\, x(s_j) = U_I[x(t)], \qquad \underset{s_j \in I}{\text{G.L.B.}}\, x(s_j) = L_I[x(t)],$$

with probability 1.

We have seen that to each I correspond sequences $\{t_j'\}$, $\{t_j''\}$, such that

$$(2.2.2) \qquad \underset{j\geqq 1}{\text{L.U.B.}}\, x(t_j') = U_I[x(t)], \qquad \underset{j\geqq 1}{\text{G.L.B.}}\, x(t_j'') = L_I[x(t)],$$

with probability 1. Let $\{s_j\}$ be any sequence including all these t_j', t_j'', for every I with rational end points. Then evidently (2.2.1) is true, with probability 1, if I is such an interval. Let I be any open t-interval, and let $I_1, I_2, \cdots$

† We define $\arctan \pm \infty$ as $\pm \pi/2$.

be intervals with rational end points, such that $I_1 \subset I_2 \subset \cdots$ and that $\sum_1^\infty I_j = I$. Then

$$\text{(2.2.3)} \qquad \underset{s_j \in I_n}{\text{L.U.B.}}\ x(s_j) = U_{I_n}[x(t)],$$

with probability 1. Let $\{\sigma_j\}$ be a sequence of t-values, including the s_j. This inclusion implies that

$$\text{(2.2.4)} \qquad \underset{s_j \in I_n}{\text{L.U.B.}}\ x(s_j) = U_{I_n}[x(t)] = \underset{\sigma_j \in I_n}{\text{L.U.B.}}\ x(\sigma_j)$$

with probability 1. Let the $\{\sigma_j\}$ also be so inclusive that

$$\text{(2.2.5)} \qquad U_I[x(t)] = \underset{\sigma_j \in I}{\text{L.U.B.}}\ x(\sigma_j)$$

with probability 1. Then if $n \to \infty$ in (2.2.4), we find that

$$\text{(2.2.6)} \qquad \underset{s_j \in I}{\text{L.U.B.}}\ x(s_j) = \underset{\sigma_j \in I}{\text{L.U.B.}}\ x(\sigma_j)$$

with probability 1, and (2.2.5), (2.2.6) taken together imply the first part of (2.2.1). The second part is proved in the same way.

LEMMA 2.3. *If a sequence $\{s_j\}$ is chosen as described in Theorem* 2.2, *and if $x^*(t)$, $x_*(t)$ are the upper and lower limiting functions of $x(t)$ on $\{s_j\}$, as defined in* (2.1.3), *then for each value of t,*

$$\text{(2.3.1)} \qquad x_*(t) \leqq x(t) \leqq x^*(t)$$

with probability 1.

For if $x^*(s_0) < x(s_0)$ for some t-value s_0, on an Ω^*-set of positive measure, and if I is an open interval containing s_0,

$$\text{(2.3.2)} \qquad U_I[x(t)] = \underset{j \geq 1}{\text{L.U.B.}}\ x(s_j) < x(s_0)$$

on an Ω^*-set of positive measure, if I is sufficiently small, contradicting the definition of $U_I[x(t)]$. Thus the second half of the inequality is proved, and the first is proved in the same way.

COROLLARY. *For each value of t, $x_*(t) < \infty$, $x^*(t) > -\infty$, with probability* 1.

It has not been necessary, as yet, to utilize a subspace Ω of Ω^*, rather than Ω^* itself. As Theorem 2.1 shows, such a transition is necessary, in discussing quasi-separable processes.

THEOREM 2.4. *Whatever the given P^*-measure, there is a space Ω of a quasi-separable process. In some cases it is necessary to allow infinite-valued functions in Ω.*

This theorem states that we can find a space on which the generalized upper and lower bounds become actual upper and lower bounds. It will be utilized to show that a regularity property of $x(t)$ known to be true when t is restricted to denumerable sets implies the same type of regularity for all t, for the functions of a suitably chosen space Ω. (Compare Theorems 2.7, 2.8, 2.9.)

Proof of Theorem 2.4. Let $\{s_j\}$ be a sequence as described in Theorem 2.2. Define $x_*(t)$, $x^*(t)$ as in (2.1.3). Let Ω be the space of all $x(t)$ with

$$x_*(t) \leqq x(t) \leqq x^*(t) \tag{2.4.1}$$

identically in t. If $x_*(t) = +\infty$, this means that $x(t) = +\infty$, and if $x^*(t) = -\infty$, this means that $x(t) = -\infty$. To prove the theorem, we need only show that $\overline{P}^*(\Omega) = 1$. Let $\Gamma = \sum_1^\infty \Gamma_n$ be any denumerable sum of neighborhoods with $\Gamma \supset \Omega$. It will be sufficient to show that $P^*(\Gamma) = 1$. Let $t_1, t_2, \cdots$ be the t-values used in defining the Γ_j. Let Ω_1 be the space of all $x(t)$ satisfying (2.4.1) for $t = t_1, t_2, \cdots$. Then by Lemma 2.3, $P^*(\Omega_1) = 1$, so it will be sufficient to prove that $\Gamma \supset \Omega_1$. Let $\xi(t)$ be any element in Ω_1. We show that $\xi(t) \in \Gamma$. We have, since $\xi(t) \in \Omega_1$,

$$\xi_*(t_j) \leqq \xi(t_j) \leqq \xi^*(t_j), \qquad j = 1, 2, \cdots. \tag{2.4.2}$$

Now there is an $x_0(t)$ in Ω such that $x_0(t_j) = \xi(t_j)$, $j \geqq 1$: we need only define $x_0(t)$ so that

$$\xi_*(t) \leqq x_0(t) \leqq \xi^*(t), \qquad t \notin \{t_j\}. \tag{2.4.3}$$

Then $x_0(t) \in \Omega \subset \Gamma$ so $\xi(t)$ is in Γ (because $\xi(t_j) = x(t_j)$, $j \geqq 1$), as was to be proved. An example will be given below to show that the introduction of infinite-valued functions may be essential in some cases. Evidently it must be known that $x^*(t)$, $-x_*(t)$ are (simultaneously for all t) not equal to $+\infty$, with probability 1, if infinite-valued functions are to be avoided.

We pass to the discussion of measurable processes. A process with space Ω is said to be measurable if the function $x_s(\omega)$ $(= x(s))$ is measurable as a function of the two variables s, ω, with $\omega \in \Omega$.†

THEOREM 2.5. *Let P^*-measure be such that there is a measurable process. Then there is a space Ω of a quasi-separable measurable process. In some cases it is necessary to allow infinite-valued functions in Ω.*

Define the function $\phi_n(t)$ by

$$\phi_n(t) = k2^{-n}, \quad \text{if} \quad (k-1)2^{-n} < t \leqq k2^{-n}, \qquad k = 0, \pm 1, \cdots, \quad n = 1, 2, \cdots. \tag{2.5.1}$$

† Cf. Doob [1, p. 113].

Then since there is a measurable process with the given P^*-measure, there is a constant c, and a sequence of integers $\{a_n\}$, such that for t fixed, not in some set e of Lebesgue measure 0,

$$\lim_{n\to\infty} x[c + \phi_{a_n}(t)] = x(c + t), \tag{2.5.2}$$

with probability 1.† Let $\{\sigma_j\}$ be the sequence of all numbers $\{c+k2^{-n}\}$, $k=0, \pm 1, \cdots, n=1, 2, \cdots$. Define $x^{**}(t)$, $x_{**}(t)$ by

$$x^{**}(t) = \limsup_{n\to\infty} x[c + \phi_{a_n}(t)], \qquad x_{**}(t) = \liminf_{n\to\infty} x[c + \phi_{a_n}(t)]. \tag{2.5.3}$$

Then for each t not in e, $x_{**}(t) = x(t) = x^{**}(t)$, with probability 1. Let $\{s_j\}$ be a sequence of numbers with the properties described in Theorem 2.2. We can suppose that the sequence $\{s_j\}$ includes the $\{\sigma_j\}$, since any sequence including the sequence $\{s_j\}$ also has the properties described in Theorem 2.2. Defining $x_*(t)$, $x^*(t)$ as in (2.1.3),

$$x_*(t) \leqq x_{**}(t) \leqq x^{**}(t) \leqq x^*(t), \tag{2.5.4}$$

and for each t, we have seen that

$$x_*(t) \leqq x(t) \leqq x^*(t), \tag{2.5.5}$$

with probability 1. Let Ω be the set of all functions $x(t)$ satisfying

$$x_{**}(t) \leqq x(t) \leqq x^{**}(t) \tag{2.5.6}$$

for $t \notin e$, and (2.5.5) for $t \in e$. (We are allowing $x(t)$ to take on infinite values.) Then if $x(t) \in \Omega$, $x(t)$ satisfies (2.5.5) identically in t. This means, according to the definition of quasi-separability, that if $\overline{P}^*(\Omega) = 1$, Ω is the space of a quasi-separable process. To prove the theorem, it will therefore be sufficient to prove that $\overline{P}^*(\Omega) = 1$ and that the function $x_s(\omega)$ is a measurable function of (s, ω), $(\omega \in \Omega)$. Let $\Gamma = \sum_1^\infty \Gamma_n$ be any denumerable sum of neighborhoods with $\Gamma \supset \Omega$. To show that $\overline{P}^*(\Omega) = 1$, we shall show that $P^*(\Gamma) = 1$. Let $t_1, t_2, \cdots$ be the t-values used in defining the neighborhoods. Let Ω be the space of all $x(t)$ such that (2.5.6) is true for the $t_j \notin e$, and (2.5.5) for the $t_j \in e$. Then $P^*(\Omega_1) = 1$, so it will be sufficient to show that $\Gamma \supset \Omega_1$. Let $\xi(t)$ be any function in Ω_1:

$$\begin{aligned} &\xi_{**}(t_j) \leqq \xi(t_j) \leqq \xi^{**}(t_j), && t_j \notin e, \\ &\xi_*(t_j) \leqq \xi(t_j) \leqq \xi^*(t_j), && t_j \in e. \end{aligned} \tag{2.5.7}$$

Now there is an $x_0(t)$ in Ω such that $x_0(t_j) = \xi(t_j)$, $j \geqq 1$: we need only have $x_0(t)$ satisfy

† Doob [1, p. 115].

$$\xi_{**}(t) \leqq x_0(t) \leqq \xi^{**}(t), \qquad t \notin \{t_j\}. \tag{2.5.8}$$

Then $x_0(t) \in \Omega \subset \Gamma$. This means that $\xi(t) \in \Gamma$ (because $x_0(t_j) = \xi(t_j), j \geqq 1$). Then $\Gamma \supset \Omega_1$: $P^*(\Gamma) = 1$. To complete the proof we must show that $x_s(\omega)$ is (s, ω)-measurable ($\omega \in \Omega$). It is readily verified that $x[c + \phi_{a_n}(s)]$, which is $x_\sigma(\omega)$ with $\sigma = c + \phi_{a_n}(s)$ is (s, ω)-measurable, and it follows that $x_s^{**}(\omega)$, $x_{s**}(\omega)$ (defined in terms of $x^{**}(t)$, $x_{**}(t)$ as $x_s(\omega)$ is in terms of $x(t)$) are (s, ω)-measurable. Then since if $s \notin e$, and if $\omega \in \Omega$ (rewriting (2.5.6))

$$x_{s**}(\omega) \leqq x_s(\omega) \leqq x_s^{**}(\omega),$$

and since the extremes are equal ($s \notin e$) for almost all $\omega \in \Omega$, the extremes are equal for almost all (s, ω). Then $x_s(\omega) = x_s^{**}(\omega) = x_{s**}(\omega)$ for almost all (s, ω) $(\omega \in \Omega)$, and so $x_s(\omega)$ is (s, ω)-measurable ($\omega \in \Omega$), as was to be proved.

COROLLARY. *The hypotheses of Theorem* 2.5 *imply the following stronger conclusion. There is a space* Ω *of a stochastic process, and a denumerable everywhere dense t-set* $\{s_j\}$ *such that, defining* $x_*(t)$, $x^*(t)$ *by* (2.1.3), *and* $y_*(t)$, $y^*(t)$ *by*

$$y_*(t) = \liminf_{s_j \downarrow t} x(s_j), \qquad y^*(t) = \limsup_{s_j \downarrow t} x(s_j), \tag{2.5.9}$$

if $x(t) \in \Omega$,

$$x_*(t) \leqq x(t) \leqq x^*(t) \tag{2.5.10}$$

for all t, and, except for a t-set e_1 *(independent of the function* $x(t)$*) of Lebesgue measure* 0,

$$y_*(t) \leqq x(t) \leqq y^*(t). \tag{2.5.11}$$

The set e_1 *contains no point* t' *with the property that whenever* $\{t_n\}$ *is a sequence converging to* t' *from above:* $t_n \downarrow t'$, *then*

$$\liminf_{n \to \infty} x(t_n) \leqq x(t') \leqq \limsup_{n \to \infty} x(t_n) \tag{2.5.12}$$

with probability 1.

The proof of Theorem 2.5 actually proves most of the corollary. Using the notation of that proof, we have

$$y_*(t) \leqq x_{**}(t) \leqq x^{**}(t) \leqq y^*(t).$$

This means that, to prove the corollary, we need only modify the definition of Ω, as follows. The space Ω is now to include all functions $x(t)$ satisfying (2.5.6) for all $t \notin e$, as before, and also for $t = t' \in e$, if t' is as described in the statement of the corollary; otherwise $x(t)$ is to satisfy (2.5.5) for $t \in e$. The proof then goes through essentially as before.

The following example of a stochastic process illustrates the application of the preceding theorem. We shall restrict t to be non-negative. Define P^*-measure as follows. Let $x(0)=0$ with probability 1. The chance variable $x(t+l)-x(t)$ $(l>0)$ has a distribution taking on only integral values, and

$$P^*\{x(t+l)-x(t)=\nu\}=\frac{e^{-l}l^\nu}{\nu!}, \qquad \nu=0,1,2,\cdots.$$

If $0\leqq t_0<\cdots<t_n$, the chance variables

$$x(t_1)-x(t_0),\cdots,x(t_n)-x(t_{n-1})$$

are to be mutually independent. According to Theorem 2.4, there is a space Ω of a quasi-separable process. Let $\{s_j\}$ be the denumerable set involved in the definition of quasi-separability, so that (2.1.4) is true for all t, if $x(t)\in\Omega$. We can assume that $t=0\in\{s_j\}$. Evidently $x(t)$, considering t only in $\{s_j\}$ is monotone non-decreasing, starting from 0, and increasing only by jumps of unit magnitude, with probability 1. Then except for an Ω-set of P-measure 0, the functions in Ω are monotone non-decreasing, and continuous except for jumps of magnitude 1, these jumps being the only points of increase. It is easily verified that Ω is the space of a measurable process.† According to the corollary to Theorem 2.5, there is even a space Ω_1 of a stochastic process all of whose functions are continuous on the right. In fact the set e_1 of that corollary is empty, because for any t, $t_n\downarrow t$ implies that $x(t_n)\to x(t)$ with probability 1. In this case infinite-valued functions need not be introduced, but we shall use this example to derive another, in which the introduction of infinite-valued functions is essential, in discussing quasi-separable processes. For each $x(t)\in\Omega_1$, as defined above, we define a function $\xi(t)$, to be 0 at $t=t_0$ if $x(t)$ has a jump there, and otherwise to be $1/|t_1-t_0|$, where t_1 is the closest point to t_0 where there is a jump. The probability relations for the $\xi(t)$ are derived from those of the $x(t)$. For each t, $\xi(t)=\xi_t(\omega)$ is a P-measurable function, and the one-parameter family of chance variables $\{\xi_t(\omega)\}$ determines a new stochastic process. The particular functions $\xi(t)$ we have defined above form a space Ω_ξ, which can be shown to be of outer measure 1 in terms of the new probability measure,‡ but we shall not need this fact. Evidently if $\xi(t)\in\Omega_\xi$, $\xi(t)$ is continuous except at an isolated denumerable set. If $\{s_j\}$ is any everywhere dense denumerable set, and if $\xi_*(t)$, $\xi^*(t)$ are defined as usual, $\xi_*(t)=+\infty$ at some point in any given interval of length l with probability $(1-e^{-l})$. Thus

$$P\left\{\max_{0\leqq t<\infty}\xi_*(t)=+\infty\right\}=1$$

† This fact also follows from Theoren. 2.5 (ii) of Doob [1].

‡ Ambrose-Doob [3].

and the introduction of infinite-valued functions is necessary to obtain a quasi-separable process for the $\xi(t)$. The space Ω_ξ is not that of a quasi-separable process, but it can be shown that if each $\xi(t)$ in Ω_ξ is altered to $+\infty$ where it vanishes, the resulting $\xi(t)$ are the functions of a quasi-separable process.

THEOREM 2.6. *Suppose that there is a t-set E such that whenever $t_n \downarrow t \in E$ $x(t_n)$ approaches a limit in probability. Then there is a denumerable set $D \subset E$ such that if $t \in E-D$, $x(t+h) \to x(t)$ in probability as $h \to 0$.*

Define a metric on the space of P^*-measurable functions $f(\omega)$, by defining the distance between f_1 and f_2 to be

$$d(f_1, f_2) = \int |\arctan f_1 - \arctan f_2| \, dP,$$

so that $d(f, f_n) \to 0$ means $f_n \to f$ in probability, and conversely, if f is finite-valued. Then for each t, the ω-function $x_t(\omega)$ becomes a point of this space, and as t varies we obtain a function of t. This function of t, $\Phi(t)$, has the property that if $t \in E$, $\lim_{t_n \downarrow t} \Phi(t_n)$ exists. Then if $t \in E$, $\lim_{t' \downarrow t} \Phi(t')$ exists. It is well known that a function with this property can have at most a denumerable number of discontinuities on E, and this set of discontinuities is the D of the theorem.

THEOREM 2.7. *Suppose that there is a t-set E such that whenever $t_n \downarrow t \in E$ $\lim_{n \to \infty} x(t_n)$ exists, with probability* 1. *Then there is a denumerable set $D' \subset E$, such that if $t \in E-D'$, $t_n \to t$ implies that $x(t_n) \to x(t)$ with probability* 1. *If Ω is the space of a quasi-separable process, each $x(t)$ in Ω is continuous almost everywhere on E with probability* 1, *and if $t_0 \in E-D'$, $x(t)$ is continuous at t_0, with probability* 1.†

Since convergence almost everywhere on Ω^* implies convergence in measure (that is, in probability) on Ω^*, Theorem 2.6 implies that there is a denumerable t-set $D \subset E$, such that $t \in E-D$, $t_n \downarrow t$ implies that $x(t_n) \to x(t)$ in probability, and since $\lim_{n \to \infty} x(t_n)$ exists with probability 1, this limit must be $x(t)$. Let S be an everywhere dense denumerable t-set. The hypothesis of Theorem 2.7 implies that if $t \in E-D$, $\lim_{t' \downarrow t} x(t')$ exists with probability 1, if t' is restricted to lie in S.‡ Let ν be any positive integer. Then if $t_0 \in E-D$, there is an open interval $I(t_0)$ with t_0 as left-hand end point, such that the oscillation of $x(t)$ ($t \in S$) on $I(t_0)$ is not greater than $1/\nu$, neglecting an Ω^*-set of measure not greater than $1/\nu$. The intervals $I(t)$ cover E except possibly

† Both in this theorem and in the preceding one, convergence to an infinite limit can be allowed. The definition of convergence in probability to an infinite limit is made in an obvious way.

‡ Doob [1, p. 111].

for an at most denumerable set $D^{(\nu)}$: for the $I(t)$ with $t \in E$, and t not covered, are nonoverlapping intervals, and so are at most denumerable. Moreover if $t_0 \in E - D^{(\nu)}$, the oscillation of $x(t)$ $(t \in S)$ at t_0 is not greater than $1/\nu$, neglecting a possible Ω^*-set (depending on t_0) of measure not greater than $1/\nu$. Then if $t_0 \in E - D_1$, where $D_1 = \sum_1^\infty D^{(\nu)}$, the oscillation of $x(t)$ $(t \in S)$ is 0, neglecting a possible Ω^*-set of measure 0. This means that if $t_0 \in E - D'$, where $D' = D + D_1$, $\lim_{t \to t_0} x(t) = x(t_0)$ $(t \in S)$, with probability 1. Now let Ω be the space of a quasi-separable process (Theorem 2.4 states that there always is such a space), and let S be a set of t-values such that (2.1.4) is satisfied for all t. It follows that if $t \in E - D'$,† $x(t)$ (in Ω) is continuous at t, with probability 1. This means, in particular, that if $t_n \to t \in E - D'$, then $x(t_n) \to x(t)$ with probability 1 in Ω, and therefore in Ω^*. The theorem will be completely proved when it is shown that $x(t)$ (in Ω) is continuous almost everywhere in E, with probability 1. To show this, let $x_s^*(\omega)$ $[x_{s*}(\omega)]$ be the function of ω: $x(t)$ which for $\omega = \omega_0$: $x_0(t)$ takes on the value $x_0^*(s)$ $[x_{0*}(s)]$. We can then write $x_s^*(\omega) = x^*(s)$, $x_{s*}(\omega) = x_*(s)$. The functions $x_s^*(\omega)$, $x_{s*}(\omega)$ are easily seen to be measurable functions of (s, ω) for $\omega \in \Omega$, and for $s \in E - D'$ are equal almost everywhere on Ω. Then by Fubini's theorem, they are equal almost everywhere on $E - D'$ (that is, almost everywhere on E) for almost all ω: $x(t)$. The equality of $x^*(t_0)$, $x_*(t_0)$ means the continuity of $x(t)$ at t_0; so the theorem is now completely proved.

THEOREM 2.8. *Suppose that P^*-measure has the property that, whatever the denumerable set S, almost every $x(t)$ in Ω^* has the property that* $\lim_{t \downarrow t_0} x(t)$ $(t \in S)$ *exists for all t_0. Let D_1 (D_2) be the t-set on which it is not true that $t_n \downarrow t$ $(t_n \uparrow t)$ implies that $x(t_n) \to x(t)$ with probability* 1. *Then D_1, D_2 are denumerable. Let Ω be the space of a quasi-separable process. Then for almost all $x(t)$ in Ω,* $\lim_{h \downarrow 0} x(t+h)$ *exists, for all t, and $x(t)$ is continuous except perhaps for a denumerable number of discontinuities (varying with the function $x(t)$ in Ω). There is a quasi-separable process with space Ω_1, for which, in addition to the above, $x(t)$ in Ω_1 is continuous on the right: $x(t+0) = x(t)$ for all $t \notin D_1$.*

The hypotheses imply that if $t_n \downarrow t$, $\lim_{n \to \infty} x(t_n)$ exists with probability 1. Then according to Theorem 2.7, D_1 and D_2 are denumerable. Let Ω be the space of a quasi-separable process, and let S be a sequence of t-values such that (2.1.4) is satisfied for all t. Then since, according to the hypotheses, $\lim_{h \downarrow 0} x^*(t+h)$, $\lim_{h \downarrow 0} x_*(t+h)$ exist and are equal for all t, with probability 1, it follows that if $x(t) \in \Omega$ and is not in an exceptional ω-set of probability 0, $\lim_{h \downarrow 0} x(t+h)$ exists, for all t. A function with this property has at most a denumerable number of discontinuities. According to the corollary to Theorem 2.5 there must be a space Ω_1, of a stochastic process, that is, with

† The set D' depends on S.

$\overline{P}^*(\Omega_1)=1$, all of whose elements are functions continuous on the right if $t \notin D_1$.

This theorem, which sounds somewhat clumsy, has been phrased with the applications in mind. It is applicable, for example, to any differential process: one in which if $t_0<t_1<\cdots<t_n$, the chance variables

$$x_{t_1}-x_{t_0},\cdots,x_{t_n}-x_{t_{n-1}}$$

are mutually independent.† The first example given above was one of such a process.

THEOREM 2.9. *Suppose that P^*-measure has the property that, whatever the denumerable t-set S, almost every $x(t)$ in Ω^* coincides on S with an everywhere continuous function $f_\omega(t)$. Then there is a space Ω of a (quasi-separable) process, whose functions $x(t)$ are everywhere continuous.*

According to Theorem 2.4, there is a quasi-separable process. Evidently almost all the functions involved in this process are everywhere continuous, so the continuous functions themselves can be taken as the functions of a process.

In many applications, the following situation arises. There is given a P^*-measure, and a one-parameter family of P^*-measurable functions $\{f_t(\omega)\}$, $-\infty<t<+\infty$. Let Ω_y^* be the space of all functions $y(t)$. We define a P^*-measure on Ω_y^* by setting

$$P_y^*\{y(t_j)<k_j, j=1,\cdots,n\}=P^*\{f_{t_j}(\omega)<k_j, j=1,\cdots,n\}.$$

In other words, we use the family of chance variables $\{f_t(\omega)\}$ to define a new stochastic process. The $y(t)$ of Ω_y^* are the elements of a stochastic process whose measure relations are those given by the $x(t)$. It is frequently desirable to restrict the $x(t)$ to some space Ω (to obtain a quasi-separable or measurable process) and to consider only the $y(t)$ given by

$$y(t)=f_t(\omega), \qquad \omega \in \Omega, \ -\infty<t<\infty.$$

It is known that the space Ω_y of these $y(t)$ has outer P_y^*-measure 1, and so is the space of a stochastic process,‡ but it is also desirable to know that this process has further properties, such as quasi-separability or measurability. In the application to be made in the next section, $f_t(\omega)$ is only determined, for each t, up to an ω-set of measure 0, and this is the usual case. We can then vary Ω_y by changing $f_t(\omega)$ for each t on an ω-set of measure 0. We shall show that by doing this we can (i) *make Ω_y the space of a quasi-separable process, or*

† Cf. Doob [1, pp. 134–137]; the theorems of the present paper make possible considerable simplification of this paper.

‡ Ambrose-Doob [3].

(ii) *(if the P_y^*-measure is such that there is a measurable process) we can make Ω_y the space of a quasi-separable measurable process.*

Proof of (i). Let $\{s_n\}$ be a t-set with the properties described in Theorem 2.2, using P_y^*-measure. Define $f_{t*}(\omega), f_t^*(\omega)$ by

$$f_{t*}(\omega) = \lim_{|I|\to 0} \underset{s_j \in I}{\text{G.L.B.}} f_{s_j}(\omega), \qquad f_t^*(\omega) = \lim_{|I|\to 0} \underset{s_j \in I}{\text{L.U.B.}} f_{s_j}(\omega),$$

where I is an open interval, containing t, of length $|I|$. Then according to Lemma 2.3, for each t,

$$f_{t*}(\omega) \leqq f_t(\omega) \leqq f_t^*(\omega)$$

almost everywhere on Ω. There is a subspace $\Omega_0 \subset \Omega$ of P-measure 1, such that these inequalities are satisfied for all $\omega \in \Omega_0$ if $t \in \{s_j\}$. If $f_{t_0}(\omega_0)$ does not satisfy this inequality for $\omega_0 \in \Omega_0$, redefine $f_{t_0}(\omega_0)$ as any number between $f_{t_0*}(\omega_0)$ and $f_{t_0}^*(\omega_0)$. We thus change $f_{t_0}(\omega)$ on at most an ω-set of measure 0, and incidentally may introduce some infinite values. The new function $g_t(\omega)$ has the properties

$$g_{s_j}(\omega) = f_{s_j}(\omega), \qquad j \geqq 1, \ \omega \in \Omega_0,$$
$$f_{t*}(\omega) = g_{t*}(\omega) \leqq g_t(\omega) \leqq g_t^*(\omega) = f_t^*(\omega).$$

Let $g_t(\omega) = 0$ if $\omega \in \Omega - \Omega_0$. Then the functions $y(t)$: $y(t) = g_t(\omega)$, $\omega \in \Omega$, determine a space Ω_y of the desired quasi-separable process.

The proof of (ii) is along the same lines, and will be omitted. The result can be extended in the same way that Theorem 2.5 was extended in its corollary.

3. **One-parameter families of chance variables with the property $\mathcal{E}$.** In this section, we shall apply the results of §2 to the special case in which the chance variables $\{x_t(\omega)\}$ defined in §2 have the property $\mathcal{E}$.† An important example of the kind of process being considered here is the following. Let Ex_t exist for all $t \geqq 0$, and suppose that $Ex_t \equiv 0$; suppose also that if $0 \leqq t_0 < \cdots < t_n$, the chance variables

$$x_{t_1} - x_{t_0}, \cdots, x_{t_n} - x_{t_{n-1}}$$

are mutually independent. The chance variables $\{x_t(\omega)\}$, $t \geqq 0$, then have the property $\mathcal{E}$ if x_0 is set identically 0 at $t = 0$. In this case, the qualitative characteristics of the random function x_t, as given by P. Lévy [6, 7], are well known: x_t can be considered continuous except for discontinuities of the first kind (jumps), or, in the terminology of the present paper, there is a space Ω

† Ville [9, pp. 111–130] has discussed families of non-negative chance variables, depending on the parameter t ranging from 0 to $+\infty$, with the property $\mathcal{E}$. His discussion of the meaning of a continuous process, and of generalized upper bounds is somewhat obscure.

of a stochastic process with the given P^*-measure, whose functions $x(t)$ are continuous except for jumps.† It will be seen below that much of this regularity is true in the more general case now under discussion.

Suppose that the chance variables $\{x_t(\omega)\}$ have the property $\mathcal{E}$. Let $t_n \to t$ and suppose that the sequence $\{t_n\}$ is monotone increasing. Then the chance variables $x_1, x_2, \cdots, x'$, with $x_n = x(t_n)$, $x' = x(t)$ satisfy the hypotheses of Theorem 1.4, so $\lim_{n\to\infty} x(t_n) = x_l(t)$ exists, with probability 1, and the variables $x(t_1), x(t_2), \cdots, x_l(t), x(t)$ have the property $\mathcal{E}$. The limit $x_l(t)$ is independent of the sequence $\{t_n\}$ (neglecting Ω^*-sets of zero P^*-measure), since two sequences can be combined into a single one, and evidently $t_n \uparrow t$ implies that $x(t_n) \to x_l(t)$ with probability 1, even if the sequence $\{t_n\}$ is not monotone. Similarly, using Theorem 1.2, $t_n \downarrow t$ implies that $\lim_{n\to\infty} x(t_n) = x_r(t)$ exists, with probability 1, and is independent of the sequence $\{t_n\}$, neglecting Ω^*-sets of zero P^*-measure. By Theorem 2.7, $x_l(t) = x_r(t) = x(t)$ with probability 1, if t is not in some set D which is at most denumerable. It also follows from Theorem 1.2 that there is a chance variable $x(-\infty)$ such that if $t_n \to -\infty$ then $x(t_n) \to x(-\infty)$ with probability 1. The family of chance variables $\{x(t)\} = \{x_t(\omega)\}$ augmented by the newly defined $x_l(t)$, $x_r(t)$, $x(-\infty)$, and ordered in the natural way, has the property $\mathcal{E}$. By the corollary to Theorem 0.2, $E|x(t)|$ is a monotone non-decreasing function of t. If $E|x(t)|$ is a bounded function, Theorem 1.3 shows that there is a chance variable $x(+\infty)$ such that if $t_n \to +\infty$, $x(t_n) \to x(+\infty)$, with probability 1. Theorem 0.3 states that the chance variables $x_t(\omega)$, for t in any set bounded on the right, are uniformly integrable.

THEOREM 3.1. *Let* $\{x_t\}$, $-\infty < t < +\infty$, *be chance variables with the property* $\mathcal{E}$. *Define* Λ, M *by*

$$\Lambda = \{U_I[x_t] > k\}, \qquad \mathrm{M} = \{L_I[x_t] < k\},$$

where I *is the interval* $a \leqq t \leqq b$. *Then if* $s \geqq b$,

$$(3.1.1) \qquad \int_\Lambda x_s dP \geqq kP(\Lambda), \qquad \int_{\mathrm{M}} x_s dP \leqq kP(\mathrm{M}).\ddagger$$

If $E|x_b \log |x_b|| < \infty$, $E\{U_I[|x_t|]\} < \infty$.

We have shown that

$$U_I[x_t] = \underset{j\geqq 1}{\text{L.U.B.}}\ x_{s_j}, \qquad L_I[x_t] = \underset{j\geqq 1}{\text{G.L.B.}}\ x_{s_j}$$

† Cf. Doob [1, pp. 134–135], and also [2] where the above specific case, involving Ex_t is discussed. The hypothesis of the existence of Ex_t is unnecessary to obtain the above regularity properties of x_t; it is the independence condition that is essential.

‡ Cf. Ville [9, pp. 113–119].

for a suitably chosen sequence s_j in the interval I. By Theorem 1.1, if

$$\Lambda_n = \left[\text{L.U.B.}_{j\leqq n}\ x_{s_j} \geqq k + 1/n\right],$$

it follows that

$$\int_{\Lambda_n} x_s dP \geqq kP(\Lambda_n), \qquad s \geqq b.$$

If n becomes infinite, this becomes the first inequality in (3.1.1), and the second is proved in the same way. If $E\{|x_b| \log |x_b|\} < \infty$, there is, according to Theorem 1.5, a constant K, depending only on this expectation, such that $E\{\text{L.U.B.}_{j\leqq n} |x_{s_j}|\} \leqq K$. If $n \to \infty$ we find that $E\{U_I|x_t|\} \leqq K$, as was to be proved.

If the chance variables $\{x_t(\omega)\}$ have the property $\mathcal{E}$, we have seen that $t_n \to t$ implies that $x(t_n) \to x(t)$ with probability 1, unless t is in an exceptional set which is at most denumerable. There is then a measurable process.† According to Theorem 2.5, there is a quasi-separable measurable process. By Theorem 3.1, if I is any finite interval, $U_I[|x(t)|] < \infty$, with probability 1. Then it is unnecessary to use infinite-valued functions in discussing quasi-separable processes. Let $D_1 \subset D$ be the (at most denumerable) set of t-values where $x(t)$, $x_\tau(t)$ are different on Ω^*-sets of positive measure. The main result of this section is that there is a space Ω of a stochastic process with the given probability measure whose elements are continuous except possibly on a denumerable t-set, and everywhere continuous on the right, except perhaps at points of D_1 (the limit $\lim_{t' \downarrow t} x(t')$ existing even on D_1). In order to prove this, we shall need several preliminary results.

LEMMA 3.2. *Let $u_\alpha(\omega)$ be a measurable function of ω, where ω varies on some space on which a measure function is defined. The subscript α ranges through the ordinals of the first and second classes. Suppose that for each ω if $\alpha_2 > \alpha_1$, $u_{\alpha_2}(\omega) > u_{\alpha_1}(\omega)$ unless $u_\beta(\omega) = u_{\alpha_1}(\omega)$ for all $\beta > \alpha$. Let $u(\omega) = \text{L.U.B.}_{\alpha \geqq 1} u_\alpha(\omega)$. Then there is a subscript ν such that $u_\nu(\omega) = u(\omega)$ for almost all ω.*

We shall only need this lemma if the $\{u_\alpha(\omega)\}$ are uniformly bounded and if the ω-measure is finite-valued, so only this case will be proved, but the extension to the general case can easily be done. Let ϵ be any positive number. We shall prove first that there is a subscript $\beta = \beta_\epsilon$ such that if $\gamma > \beta$, $u_\gamma - u_\beta < \epsilon$ except possibly on a set (depending on γ), of measure less than ϵ. If this is not so there are pairs (β_1, γ_1), (β_2, γ_2), $\cdots$ with $\beta_1 < \gamma_1 < \beta_2 < \gamma_2 < \cdots$ such that $\mu_{\gamma_n}(\omega) - \mu_{\beta_n}(\omega) \geqq \epsilon$ on an ω-set Λ_n of measure not less than ϵ. The set of points Λ in infinitely many Λ_n has measure at least ϵ. Now this is impossible, since

† Doob [1, p. 118].

then $u(\omega)=+\infty$ on Λ_n, and we have supposed that the u_α are uniformly bounded. There are therefore subscripts $\beta_1, \beta_2, \cdots, \beta_1<\beta_2<\cdots$ such that (we are choosing ϵ successively as $1/2, 1/3, \cdots$ in the above result) if $\gamma>\beta_n$

$$u_\gamma - u_{\beta_n} < 1/n,$$

except possibly on a set of measure less than $1/n$. The sequence $\{\mu_{\beta_n}(\omega)\}$ is monotone; let $\lim_{n\to\infty} u_{\beta_n}=u_\infty$. Let ν be any ordinal of the second class, beyond $\beta_1, \beta_2, \cdots$. Then $\mu_\nu \geqq \mu_{\beta_n}$, $n \geqq 1$, so $u_\nu \geqq u_\infty$ for almost all ω. On the other hand, $u_\nu \leqq u_{\beta_n}+1/n$, except possibly on a set of measure not greater than $1/n$. Therefore $u_\nu \leqq u_\infty$ for almost all ω. It follows that $u_\nu=u_\infty$ for almost all ω and, since ν was any ordinal of the second class beyond $\beta_1, \beta_2, \cdots$, $u_{\nu+1}=u_\infty$, for almost all ω also. Thus except for an ω-set Λ_0 of zero measure, $u_{\nu+1}=u_\infty$. This implies that if $\omega \notin \Lambda_0$, $u_\nu=u_{\nu+1}=\cdots=u$, as was to be proved.

LEMMA 3.3. *Suppose that P^*-measure has the property that for almost all t, $t_n \to t$ implies that $x(t_n)\to x(t)$ with probability 1. There is then a space Ω of a quasi-separable measurable process. Let $\phi(\omega)$ be an Ω-measurable function and let $\phi_n(\omega)=\nu/n$, where*

$$(\nu-1)/n \leqq \phi(\omega) < \nu/n, \qquad \nu = 0, \pm 1, \cdots, n = 1, 2, \cdots.$$

Then $x[s+\phi_n(\omega)]$, $x[s+\phi(\omega)]$† are measurable (s, ω)-functions $(\omega \in \Omega)$ and $x[s+\phi_n(\omega)]\to x[s+\phi(\omega)]$ almost everywhere on (s, ω)-space $(\omega \in \Omega)$.

According to Theorem 2.4, there is a quasi-separable process with the given P^*-measure. The hypotheses have been shown to imply that any quasi-separable process is measurable.‡ Let Ω be the space of a quasi-separable measurable process. To show that $x[s+\phi(\omega)]$ is (s, ω)-measurable $(\omega \in \Omega)$, we must show that for any k, $\{x[s+\phi(\omega)]>k\}$ is an (s, ω)-measurable set. Since $x_s(\omega)$ is (s, ω)-measurable $(\omega \in \Omega)$, because of the measurability of the process with space Ω, we need only show that if Λ is an (s, ω)-measurable set, the points (s, ω) with $(s+\phi(\omega), \omega) \in \tilde{\Lambda}$ constitute an (s, ω)-measurable set. Denote this image of $\tilde{\Lambda}$ by $\tilde{\Lambda}_1$. We shall show that the transformation $\tilde{\Lambda}\to\tilde{\Lambda}_1$ takes measurable sets into measurable sets. Suppose first that $\tilde{\Lambda}=I\times\Lambda$: the direct product of an s-interval $a\leqq s\leqq b$ and a measurable Ω-set Λ. Then $\tilde{\Lambda}_1$ is a simple ordinate set: the (s, ω) points "over" Λ determined by the inequalities

$$a - \phi(\omega) \leqq s \leqq b - \phi(\omega),$$

and an ordinate set of this type is known to be measurable. It follows that if Λ

† The function $x[s+\phi(\omega)]$ is the function $x_\sigma(\omega)$, where $\sigma=s+\phi(\omega)$, and $x[s+\phi_n(\omega)]$ is defined similarly.

‡ Doob [1, p. 118].

is in the Borel field $\mathcal{J}$ of sets determined by the sets which are direct products of s-intervals and Ω-measurable sets, the corresponding $\tilde{\Lambda}_1$ is (s, ω)-measurable. In particular suppose that $\tilde{\Lambda} \in \mathcal{J}$ and that $\tilde{\Lambda}$ has measure 0. Then for almost all ω_0 the s-set $\{(s, \omega_0) \in \tilde{\Lambda}\}$ is of measure 0. This same fact will certainly be true of the s-set $\{(s+\phi(\omega_0), \omega_0) \in \tilde{\Lambda}\}$ so the set $\tilde{\Lambda}_1$ will have measure 0. Since any (s, ω)-measurable set of measure 0 can be included in some set of $\mathcal{J}$ which is of measure 0, it follows that whenever $\tilde{\Lambda}$ is measurable and has measure 0, the same is true of $\tilde{\Lambda}_1$. Now the most general (s, ω)-measurable set $\tilde{\Lambda}$ differs from some set $\tilde{M}$ of $\mathcal{J}$ at most by a set of measure 0, so the corresponding $\tilde{\Lambda}_1$ will differ from $\tilde{M}_1$ by at most a set of measure 0: $\tilde{\Lambda}_1$ is measurable, as was to be proved. We have thus proved that $x[s+\phi(\omega)]$ (and therefore $x[s+\phi_n(\omega)]$) is (s, ω)-measurable ($\omega \in \Omega$). According to Theorem 2.7, if $x(t) \in \Omega$, $x(t)$ is continuous for almost all t, with probability 1. Then if ω is fixed, $x[s+\phi_n(\omega)] \to x[s+\phi(\omega)]$ for almost all s, with probability 1. This implies that $x[s+\phi_n(\omega)] \to x[s+\phi(\omega)]$ for almost all (s, ω), which is the final conclusion of the lemma.

LEMMA 3.4. *Let $u(\omega)$ be an ω-measurable function, defined on the space Ω of a stochastic process whose chance variables $\{x_t(\omega)\}$ have the property $\mathcal{E}$. Suppose that $u(\omega)$ takes on only a finite number of values: $u(\omega)=\lambda_j$ on Λ_j, $j=1, 2, \cdots$. Suppose that for each j, $\{u(\omega)=\lambda_j\}$ is an ω-set depending on ω: $x(t)$ for $t<\lambda_j$. Then the chance variables $\{y_t\}$, $y_t=x[t+u(\omega)]$ (that is, $y_t=x_\sigma(\omega)$, $\sigma=t+u(\omega)$), have the property $\mathcal{E}$, for $t>0$.*

To prove this fact, we need only show that if M is a P^*-measurable set depending on the $y(t)$ for $t=t_1, \cdots, t_\nu$, $0<t_1< \cdots <t_\nu<t_\nu+h$, then

$$\int_M y_{t_\nu+h} dP = \int_M y_{t_\nu} dP. \tag{3.4.1}$$

Now

$$\int_M y_{t_\nu+h} dP = \sum_j \int_{\Lambda_j \cdot M} y_{t_\nu+h} dP = \sum_j \int_{\Lambda_j \cdot M} x(t_\nu + h + \lambda_j) dP. \tag{3.4.2}$$

The conditions on the $y(t_i)$ determining M are, on Λ_j, conditions on $x(t_i+\lambda_j)$, $i=1, \cdots, \nu$, so $\Lambda_j \cdot M$ is a set depending on t-values for $t \leqq t_\nu+\lambda_j$. Because of the fact that

$$E[x_t, t \leqq t_\nu + \lambda_j; x_{t_\nu+\lambda_j+h}] = x_{t_\nu+\lambda_j},$$

with probability 1 (Theorem 0.1), we find that

$$\int_{\Lambda_j \cdot M} x(t_\nu + \lambda_j + h) dP = \int_{\Lambda_j \cdot M} x(t_\nu + \lambda_j) dP = \int_{\Lambda_j \cdot M} y_{t_\nu} dP. \tag{3.4.3}$$

Summing over j gives

$$\int_M y_{t_\nu+h}dP = \int_M x(t_\nu + u + h)dP = \int_M y_{t_\nu}dP, \tag{3.4.4}$$

or (3.4.1), as was to be proved.

LEMMA 3.5. *Let $u(\omega)$ be a bounded ω-measurable function, defined on the space Ω of a stochastic process whose chance variables $x_t(\omega)$ have the property $\mathcal{E}$. Suppose that for each number c, $\{u(\omega) \geqq c\}$ is an ω-set depending on ω: $x(t)$ for $t<c$. Let S be any everywhere dense denumerable set. Then* $\lim_{s \to 0} x[s+u(\omega)]$† *$(s \in S)$ exists with probability* 1.

Let $u_n(\omega) = \nu/n$ where

$$(\nu - 1)/n \leqq u(\omega) < \nu/n, \qquad \nu = 0, \pm 1, \cdots, n = 1, 2, \cdots.$$

Then $u_n(\omega)$ satisfies the hypotheses of Lemma 3.4. Let $y_n(s) = x[s+u_n(\omega)]$, $y(s) = x[s+u(\omega)]$. We show first that, for fixed n and $s>0$, if $T > s+\text{L.U.B.}_\omega u(\omega)$, then

$$E[y_n(s); x(T)] = y_n(s), \tag{3.5.1}$$

with probability 1, that is, that the ordered pair of chance variables $y_n(s)$, $x(T)$ has the property $\mathcal{E}$. Let Λ be any ω-set depending on y_n, that is, determined by some condition imposed on y_n. It must be shown that

$$\int_\Lambda y_n(s)dP = \int_\Lambda x(T)dP. \tag{3.5.2}$$

Let $\Lambda_{n,j} = \{u_n = j/n\}$. Then

$$\int_\Lambda y_n(s)dP = \sum_j \int_{\Lambda \cdot \Lambda_{n,j}} x(s + j/n)dP. \tag{3.5.3}$$

Now the ω-set $\Lambda_{n,j}$ depends only on t-values less than j/n, and on $\Lambda_{n,j}$ the condition on y_n determining Λ becomes one on $x(s+j/n)$, so $\Lambda \cdot \Lambda_{n,j}$ depends only on t-values not greater than $s+j/n$. This means that, using the property $\mathcal{E}$,

$$\int_\Lambda y_n(s)dP = \sum_j \int_{\Lambda \cdot \Lambda_{n,j}} x(T)dP = \int_\Lambda x(T)dP, \tag{3.5.4}$$

or (3.5.1), as was to be shown. Now if n varies, it follows that the $y_n(s)$ are uniformly integrable, from the proof of Theorem 0.3: in fact, the proof of that theorem did not use the full property $\mathcal{E}$, but only the fact that, for

† The function $x[s+u(\omega)]$ is the function $x_\sigma(\omega)$, where $\sigma = s+u(\omega)$.

$t \leqq t_0$, $E[x_t;\ x_{t_0}] = x_t$ with probability 1. This corresponds to the fact that $E[y_n;\ x(T)] = y_n$, $n = 1, 2, \cdots$, with probability 1.

Having these preliminary facts in mind, we proceed to prove Lemma 3.1. We enlarge S to include the rational numbers, and so that S contains the number $s_1 + s_2$ if it contains s_1 and s_2. Let δ_1, δ_2 be positive numbers, with $\delta_1 < \delta_2$. According to Lemma 3.4, the $y_n(s)$ for $s > 0$ (n fixed) have the property $\mathcal{E}$. Then it can be verified at once that the chance variables $y_n(s) - y_n(\delta_1)$, for $s > \delta_1$ (n fixed) have the property $\mathcal{E}$. Using Theorem 3.1, we find that

$$(3.5.5)\quad P\left\{\text{L.U.B.}_{\delta_1 < s < \delta_2} [y_n(s) - y_n(\delta_1)] > \epsilon\right\} \leqq \frac{1}{\epsilon}\int_\Lambda [y_n(\delta_2) - y_n(\delta_1)]\,dP,\quad s \in S,$$

where Λ is the set on the left, and

$$(3.5.6)\quad P\left\{\text{G.L.B.}_{\delta_1 < s < \delta_2} [y_n(s) - y_n(\delta_1)] < -\epsilon\right\} \leqq \frac{1}{\epsilon}\int_{\mathrm{M}} [y_n(\delta_2) - y_n(\delta_1)]\,dP,\ s \in S,$$

where M is the set on the left. Then surely

$$(3.5.7)\quad \begin{aligned} P\left\{\text{L.U.B.}_{\delta_1 < s < \delta_2} [y_n(s) - y_n(\delta_1)] > \epsilon\right\} &\leqq \frac{1}{\epsilon}\int |\, y_n(\delta_2) - y_n(\delta_1)\,|\,dP, \\ P\left\{\text{G.L.B.}_{\delta_1 < s < \delta_2} [y_n(s) - y_n(\delta_1)] < -\epsilon\right\} &\leqq \frac{1}{\epsilon}\int |\, y_n(\delta_2) - y_n(\delta_1)\,|\,dP, \end{aligned} \qquad s \in S,$$

so that

$$(3.5.8)\quad P\left\{\text{L.U.B.}_{\delta_1 < s < \delta_2} |\, y_n(s) - y_n(\delta_1)\,| > \epsilon\right\} \leqq \frac{2}{\epsilon}\int |\, y_n(\delta_2) - y_n(\delta_1)\,|\,dP,\ s \in S,$$

which implies that

$$(3.5.9)\quad P\left\{\text{L.U.B.}_{\delta_1 < s', s'' < \delta_2} |\, y_n(s') - y_n(s'')\,| > 2\epsilon\right\} \leqq \frac{2}{\epsilon}\int |\, y_n(\delta_2) - y_n(\delta_1)\,|\,dP, \qquad s', s'' \in S.$$

Now if ξ_n denotes the quantity on the left in (3.5.9),

$$\begin{aligned} \xi_n &= \sum_\nu P\left\{\text{L.U.B.}_{\delta_1 < s', s'' < \delta_2} |\, x(s' + \nu/n) - x(s'' + \nu/n)\,| > 2\epsilon;\ u_n = \nu/n\right\} \\ &= \sum_\nu P\left\{\text{L.U.B.}_{\delta_1 + \nu/n < s', s'' < \delta_2 + \nu/n} |\, x(s') - x(s'')\,| > 2\epsilon;\ u_n = \nu/n\right\}, \qquad (3.5.10) \\ &\geqq P\left\{\text{L.U.B.}_{\delta_1 + u + 1/n < s', s'' < \delta_2 + u} |\, x(s') - x(s'')\,| > 2\epsilon\right\}, \qquad s', s'' \in S. \end{aligned}$$

It is easily verified that all the sets involved are ω-measurable. When n becomes infinite, we obtain

$$(3.5.11)\quad \liminf_{n\to\infty} \xi_n \geqq P\left\{ \text{L.U.B.}_{\delta_1+u<s',s''<\delta_2+u} \mid x(s') - x(s'') \mid > 2\epsilon \right\}, \qquad s', s'' \in S.$$

It was shown in Lemma 3.3 that $y_n(s)\to y(s)$ for almost all (s, ω). Then except for some s-set e of measure 0, $y_n(s)\to y(s)$ for almost all ω. Choose δ_1, δ_2 not in e. Combining (3.5.9), (3.5.10), (3.5.11), we find that

$$(3.5.12)\quad P\left\{ \text{L.U.B.}_{\delta_1+u<s',s''<\delta_2+u} \mid x(s') - x(s'') \mid > 2\epsilon \right\} \leqq \frac{2}{\epsilon}\int \mid y(\delta_2) - y(\delta_1) \mid dP, \qquad s', s'' \in S.$$

We can integrate to the limit because the y_n have been shown to be uniformly integrable. Now let $\delta_1\to 0$ along a sequence of values remaining out of the set e. By Theorem 1.2, $y(\delta)$ approaches a limit with probability 1, say $y(0+)$, and we can integrate term by term, so that

$$(3.5.13)\quad P\left\{ \text{L.U.B.}_{u<s',s''<\delta_2+u} \mid x(s') - x(s'') \mid > 2\epsilon \right\} \leqq \frac{2}{\epsilon}\int \mid y(\delta_2) - y(0+) \mid dP, \qquad s', s'' \in S.$$

If now $\delta_2\to 0$ along the same sequence of values, the integral goes to 0. This means that the oscillation of $x(t)$ at $t=u(\omega)$ on the right (only considering the function $x(t)$ defined on S) is not greater than 2ϵ, with probability 1. Since ϵ is arbitrary, the lemma follows at once from this fact.

THEOREM 3.6. *Suppose that the chance variables $\{x_t\}$ have the property $\mathcal{E}$. Let D_1 be the (at most denumerable) set of t-values for which $x(t)$, $x_r(t)$ are not identical with probability* 1. *There is a space Ω of a quasi-separable measurable stochastic process whose elements $x(t)$ have the following properties: each element $x(t)$ is continuous except possibly at a denumerable set of t-values (which may vary with the element $x(t)$); $\lim_{h\downarrow 0} x(t+h)$ exists for all t, and the limit is $x(t)$ if $t \notin D_1$; $\lim_{h\downarrow 0} x(t+h) = x_r(t)$, $\lim_{h\uparrow 0} x(t+h) = x_l(t)$ with probability* 1, *at each t-value.*†

The results obtained in this theorem are less strong than those obtained in the case of a differential process, discussed above, in which the increments of $x(t)$ in nonoverlapping intervals are independent chance variables. In the differential case, the only discontinuities are jumps, whereas here the elements $x(t)$ can apparently have complicated discontinuities on the left (of probability 0 at any fixed t-value, however, except where $x_l(t)$, $x(t)$ are not

† The exceptional Ω-set depends on t. It will be remembered that except for a denumerable t-set, $x_l(t) = x_r(t) = x(t)$ with probability 1 (t fixed).

equal with probability 1, and there the discontinuity is a jump, with probability 1). Some such difference might have been suspected, since the property $\mathcal{E}$ is essentially unsymmetric in t.

Proof of Theorem 3.6. Let S be any everywhere dense denumerable set, and let ϵ be any positive number. Consider the t-set $E[x(t)]$ where the oscillation of $x(t)$ $(t \in S)$ is not less than ϵ. This is a closed set. The ω-set

$$(3.6.1)\quad \prod_{\nu=1}^{\infty} \sum_{|s'-s''|<1/\nu,\, a-1/\nu<s',s''<b+1/\nu} \{\,|\, x(s') - x(s'')\,| \geqq \epsilon - 1/\nu\}, \quad s', s'' \in S,$$

is the set of elements $x(t)$ for which E has at least one point in $a \leqq t \leqq b$. This set is evidently P^*-measurable. Let k be any positive integer. For each $x(t)$ define $u[x(t)]$ as the first value of t with $-k \leqq t \leqq k$ in $E[x(t)]$ at which the discontinuity on the right is greater than 0;† if there is no such point, let $u=k$. The points s of $E[x(t)]$ with $-k \leqq s < u[x(t)]$ define a set $F[x(t)]$ with the following properties: there is no sequence $s_1, s_2, \cdots$ in $F[x(t)]$ with $s_1 > s_2 > \cdots$; $F[x(t)]$ is at most denumerable. Then if $F[x(t)]$ is not empty, and is ordered from left to right, it is a well-ordered set, whose ordinal number is in the first or second class. If α is any ordinal number in the first or second class, define $u_\alpha[x(t)]$ as the αth point of $F[x(t)]$ (or $u_\alpha = u$, if there is no αth point). Then if $\beta > \alpha$, $u_\beta[x(t)] > u_\alpha[x(t)]$ unless $u_\alpha[x(t)] = u_{\alpha+1}[x(t)] = \cdots = u[x(t)]$, and for each function $x(t)$ there will be an α such that the above equalities are true. Concluding this preliminary discussion, we come to the proof of the theorem, which we shall divide into three parts.

I. Let k_1, k_2 be any numbers with $k_1 < k_2$. Let $v[x(t)]$ be the first point of $E[x(t)]$ in the interval $k_1 \leqq t \leqq k_2$, or if there is no such point let $v[x(t)] = k_2$. Then we prove that

(i) $v[x(t)]$ is a measurable function of ω,

(ii) the ω-set $v \geqq c$, for any number c, depends only on t for $t < c$, and

(iii) the oscillation of $x(t)$ $(t \in S)$ on the right at $v[x(t)]$ is 0.

To prove (i) it is sufficient to note that $k_1 \leqq v \leqq k_2$ and that the (measurable) ω-set (3.6.1) with $a = k_1$, $k_1 < b < k_2$, is the set $\{v \leqq b\}$. To prove (ii) we note first (from (3.6.1) with $a = k_1$, $b = c - \eta$) that the ω-set $\{v \leqq c - \eta\}$ $(k_1 < c < k_2)$ depends only on t-values less than $c - \eta/2$, for any $\eta > 0$. Letting $\eta \to 0$, we find that the ω-set $\{v < c\}$ and therefore also $\{v \geqq c\}$ depends only on t-values less than c. It then follows that (ii) holds for all c. Lemma 3.5 can now be applied to show that $\lim_{s \downarrow v} x(s)$ $(s \in S)$ exists with probability 1, that is, that (iii) is true.

---

† The discontinuity on the right at t_0 is defined as $\lim_{\delta \to 0}$ L.U.B.$_{t_0 < s', s'' < t_0 + \delta} |x(s') - x(s'')|$ $(s', s'' \in S)$.

II. We next prove that for each α

(i) $u_\alpha[x(t)]$ is a measurable function of ω and

(ii) $\{u_\alpha[x(t)] \geqq c\}$, for any number c, depends only on t-values for $t<c$.

The proof will be given by transfinite induction. Consider the case $\alpha=1$. Let $v[x(t)]$ be the first point of $E[x(t)]$ in the interval $-k \leqq t \leqq k$, or let $v=k$ if there is no such point. Then I shows that $v=u_1$ with probability 1, and that u_1 has the properties (i) and (ii) since v has. Suppose that (i) and (ii) are true for some α. We shall show that they are then true for $\alpha+1$. Let $w[x(t)]$ be the first point of $E[x(t)]$ in the interval $u_\alpha[x(t)]<t \leqq k$; let $w=k$, if there is no such point or if $u_\alpha=k$. We can obtain w as follows: for each rational number $r<k$ determine the ω-set $\{u_\alpha<r\}$; on this set w is the first point of $E[x(t)]$ beyond r, and not greater than k, or $w=k$ if there is no such point; if $u_\alpha=k$, $w=k$ also. This makes it obvious, using I, that $w=u_{\alpha+1}$ with probability 1. Properties (i), (ii) are true for $u_{\alpha+1}$ since they are true for w (as can be seen using I and the fact that (i), (ii) are known to be true for u_α). To finish the induction proof of (i), (ii), we must show that if $\alpha_1<\alpha_2<\cdots$, and if α is the first ordinal beyond the α_j, then (i), (ii) are true for u_α if they are supposed true for $u_{\alpha_1}, u_{\alpha_2}, \cdots$. Let $w=\lim_{j\to\infty} u_{\alpha_j}$. Then w is a P^*-measurable function, $w \leqq u$, and if $w[x(t)]<k$, $w[x(t)]$ is certainly a point of $E[x(t)]$. Moreover, since (ii) is true for $u_{\alpha_1}, u_{\alpha_2}, \cdots$, it follows easily that (ii) is true for $w[x(t)]$. Lemma 3.5 now shows that $x(t)$ has 0 oscillation on the right ($t \in S$) at w, with probability 1, so that $w[x(t)]=u_\alpha[x(t)]$, with probability 1. Therefore u_α has the properties (i) and (ii), since w has. We have thus proved that (i), (ii) hold for all α.

III. The functions $\{u_\alpha(\omega)\}$ satisfy the hypotheses of Lemma 3.2, so there is an index ν such that $u_\nu=\text{L.U.B.}_\alpha\, u_\alpha=u$ except possibly on an ω-set of measure 0. It is impossible that $u_\nu=u<k$ on an ω-set of positive measure, because u_ν is a point with 0 oscillation of $x(t)$ ($t \in S$) on the right, with probability 1 (Lemma 3.5), whereas u is by hypothesis a point where the oscillation on the right is positive, if $u<k$. Then $u_\nu=u=k$ almost everywhere on Ω^*. The indices $\alpha \leqq \nu$ are denumerable, at most, so we have proved that if an ω-set of measure 0 is excluded, $\lim_{t' \downarrow t} x(t')$ ($t' \in S$) exists for all t, with $|t| \leqq k$, where the oscillation of $x(t)$ ($t \in S$) is at least ϵ. If we apply this result letting ϵ run through the values 1, 1/2, 1/3, $\cdots$ and k through the values 1, 2, $\cdots$, we find that neglecting some ω-set of measure 0, $\lim_{t' \downarrow t} x(t')$ ($t' \in S$) exists for all t. Theorem 2.8 can now be applied to give the conclusion of the present theorem, except that part involving $\lim_{h \downarrow 0} x(t+h)$, $\lim_{h \uparrow 0} x(t+h)$ (t fixed), $\lim_{t\to\infty} x(t)$, $\lim_{t\to-\infty} x(t)$. We have seen above, in our preliminary discussion of continuous stochastic processes whose chance variables $\{x_t(\omega)\}$ have the property $\mathcal{E}$, that if $t_n \downarrow t$, $x(t_n) \to x_r(t)$ with probability 1. This implies that if S

is any denumerable set, $\lim_{t' \downarrow t} x(t') = x_r(t)$ $(t' \in S)$, with probability 1†; so if the process is quasi-separable, $\lim_{h \downarrow 0} x(t+h) = x_r(t)$, with probability 1. The other cases are treated similarly.

THEOREM 3.7. *Let P^*-measure be arbitrary, and let x be any chance variable depending on the x_t (that is, a P^*-measurable function) whose expectation exists. Let Ω be the space of a stochastic process with this P^*-measure. Then if $y(t) = E[x_s, s \leqq t; x]$,‡ these conditional expectations can be defined so that for ω fixed, in Ω, not in some set of probability 0, $y(t)$ is continuous except possibly on a denumerable set, and even on this set $\lim_{h \downarrow 0} y(t+h)$ exists; moreover $\lim_{t \to -\infty} y(t)$ exists, and $\lim_{t \to +\infty} y(t) = x$.*

The chance variables $y(t)$ have the property $\mathcal{E}$. This theorem follows readily from Theorem 3.6, using the discussion of families of P^*-measurable functions at the end of §2. The fact that the limit as $t \to +\infty$ is x follows from Theorem 1.4.

BIBLIOGRAPHY

1. J. L. Doob, *Stochastic processes depending on a continuous parameter*, these Transactions, vol. 42 (1937), pp. 107–140.

2. ———, *The law of large numbers for continuous stochastic processes*, to appear in the Duke Mathematical Journal.

3. J. L. Doob and Warren Ambrose, to appear in the Annals of Mathematics.

4. A. Khintchine, *Asymptotische Gesetze der Wahrscheinlichkeitsrechnung*, Ergebnisse der Mathematik, vol. 2, no. 4.

5. A. Kolmogoroff, *Grundbegriffe der Wahrscheinlichkeitsrechnung*, Ergebnisse der Mathematik, vol. 2, no. 3.

6. P. Lévy, *Sur les intégrales dont les éléments sont des variables aléatoires indépendantes*, Annali delle Università Toscane, Pisa, (2), vol. 3 (1934), pp. 337–366.

7. ———, *Sur un précédent mémoire de l'auteur*, ibid., (2), vol. 4 (1935), pp. 217–218.

8. ———, *Théorie de l'Addition des Variables Aléatoires*, Paris, 1937.

9. Jean Ville, *Étude Critique de la Notion de Collectif*, Paris, 1939.

10. N. Wiener, *The ergodic theorem*, Duke Mathematical Journal, vol. 5 (1935), pp. 1–18.

† Doob [1, p. 111].

‡ Each conditional expectation is only uniquely defined up to an Ω^*-set of measure 0.

UNIVERSITY OF ILLINOIS,
URBANA, ILL.

ON THE ARITHMETIC OF QUATERNIONS*

BY

GORDON PALL

1. Some fifteen references, beginning with Euler, abstracted in Dickson's *History*,† use a connection between the congruence and equation

$$1) \qquad v_1^2 + v_2^2 + v_3^2 \equiv 0 \pmod{m}, \qquad t_0^2 + t_1^2 + t_2^2 + t_3^2 = m,$$

in the case where m is a prime, to prove that every positive integer is a sum of four squares. This connection is made precise, for arbitrary odd m, in Theorem 4. Among our theorems will be found conditions for quaternions to have the same right or left divisors of given norms. An easy derivation of known relations concerning binary quadratic class-numbers and representation in $x_1^2+x_2^2+x_3^2$ concludes the article.

Most of these results were discovered during an investigation of the "rational automorphs of $x_1^2+x_2^2+x_3^2$," and many of them are used in an associated article of that name. Since these automorphs are connected more simply with Lipschitz quaternions than with those of Hurwitz in which the coordinates may be halves of odds, our results are stated for the former type of integral quaternion, although there is no difficulty in extending them to the latter. No further mention is made in this article of automorphs.‡

Notations. The letters $a, b, c, t, u, \cdots, z$ denote integral quaternions of the type $t=t_0+i_1t_1+i_2t_2+i_3t_3$, with rational integer coordinates t_i. Conjugate and norm are defined as usual: $\bar{t}=t_0-\sum i_\alpha t_\alpha$, $Nt=\bar{t}t=\sum t_i^2$. Except for the quaternion units i_α, letters with subscripts, and $d, e, \cdots, s$ denote rational integers; p denotes an odd prime, *m an odd positive integer*. Subscripts α or β range over 1, 2, 3; i, j, f, or g over 0, 1, 2, 3.

We call t *pure* (mod m) if $m \mid t_0$; *proper* (mod m) if the g.c.d. of $t_0, \cdots, t_3$, and m is 1; *proper* if the g.c.d. of $t_0, \cdots, t_3$ is 1.

German capitals represent the sets defined as follows:

---

* Presented to the Society in part April 15, 1933, under the title *On the relations between sums of three and four squares*, and in part April 13, 1940; received by the editors January 31, 1940. This paper was received by the editors of the Annals of Mathematics July 5, 1939, accepted by them, and later transferred to these Transactions.

† L. E. Dickson, *History of the Theory of Numbers*, vol. 2, chap. 8.

‡ The complications of Lipschitz's article are partly due to their use. (R. Lipschitz, Journal de Mathématique, (4), vol. 2 (1886), pp. 373–439.) Several papers by the writer, now published, owe their inception to the study of the rational automorphs, and some of their proofs now stated in terms of quaternions were originally obtained by automorphs. The associated article mentioned above will appear soon in the Annals of Mathematics.

$\mathfrak{Q}$: eight left-associate proper quaternions $\pm t$, $\pm i_\alpha t$;

$\mathfrak{E}$: the set obtained from a given proper quaternion t by an even number of sign-changes or interchanges of the coordinates t_i;

$\mathfrak{M}$: m pure quaternions of norm zero, and proportional, (mod m); that is $0, v, 2v, \cdots, (m-1)v$ (mod m), where v is pure and proper (mod m), $m \mid Nv$.

2. If t is a right divisor of x, that is, $x = ut$ in integral quaternions, then the left-associates $\pm t$, $\pm i_\alpha t$ are right divisors of x with the same norm.

THEOREM 1. *If $v = v_0 + i_1 v_1 + i_2 v_2 + i_3 v_3$ is proper* (mod m), $m \mid Nv$, *m odd and positive, then v has precisely one set $\mathfrak{Q}$ of right divisors of norm m.*

This theorem, fundamental in the arithmetic of quaternions, was proved by Lipschitz* in the case of a prime m by a modification of Euler's method of proof that every prime is a sum of four squares. Hermite† had an elegant device for proving the four square theorem, based on the fact that there is only one class of positive quaternary quadratic forms of determinant 1. The following proof of Theorem 1 is an adaptation of Hermite's method. An extension of this proof to generalized quaternions has already been published.‡

LEMMA 1. *If $x \equiv y$* (mod m), *x and y have the same right divisors of norm m.*

For if $v = ut$ and $Nt = m$, $v + zm = (u + z\bar{t})t$.

LEMMA 2. *If Nx is prime to m, v and xv have the same right divisors of norm m.*

For if $v = ut$, $xv = (xu)t$. Conversely if $xv = ut$, choose k so that $kNx \equiv 1$ (mod m); *then* $v \equiv (k\bar{x}u)t$ (mod m), and we apply Lemma 1.

LEMMA 3. *If Theorem 1 holds for every product m of $r-1$ primes or less ($r > 1$), it holds for products m of r primes.*

See the top ten lines on page 702 of the reference just cited.

LEMMA 4. *If v is proper* (mod p), *we can choose a pure quaternion x of norm prime to p, such that xv is pure* (mod p).

We assume as we may that $p \nmid v_1$, and must choose x to satisfy

$$x_1 v_1 + x_2 v_2 + x_3 v_3 \equiv 0, \qquad x_1^2 + x_2^2 + x_3^2 \not\equiv 0 \pmod{p}. \tag{2}$$

Solving (2_1) in the form $x_1 \equiv ex_2 + fx_3$, we reduce (2_2) to

$$(1 + e^2)x_2^2 + 2efx_2x_3 + (1 + f^2)x_3^2 \not\equiv 0 \pmod{p}.$$

* Loc. cit., pp. 416–420.

† C. Hermite, Journal für die reine und angewandte Mathematik, vol. 47 (1854), pp. 343–345; *Oeuvres*, vol. 1, pp. 234–237.

‡ G. Pall, Duke Mathematical Journal, vol. 4 (1938), pp. 696–704.

Since these three coefficients are not all zero (mod p), Lemma 4 follows.

By these lemmas the proof of Theorem 1 reduces to the case where m is a prime p, and v is a pure quaternion such that

$$v = i_1 - i_2 v_2 - i_3 v_3, \qquad 1 + v_2^2 + v_3^2 = pq, \qquad q \text{ an integer.} \tag{3}$$

How many quaternions t satisfy $v = ut$, $Nt = p$, or what is equivalent:

$$v\bar{t} \equiv 0, \qquad Nt = p? \tag{4}$$

Condition (4_1) expands into four linear congruences in the t_i. In view of $1+v_2^2+v_3^2 \equiv 0$ these reduce to the two congruences

$$t_0 \equiv v_3 t_2 - v_2 t_3, \qquad t_1 \equiv v_2 t_2 + v_3 t_3 \pmod{p}. \tag{5}$$

We therefore substitute in the condition $\sum t_i^2 = p$ the expressions

$$t_0 = pX_0 + v_3 X_2 - v_2 X_3, \quad t_1 = pX_1 + v_2 X_2 + v_3 X_3, \quad t_2 = X_2, \quad t_3 = X_3, \tag{6}$$

where the X_i are integers, and, on dividing through by p, obtain

$$p(X_0^2 + X_1^2) + 2v_3(X_0X_1 + X_2X_3) - 2v_2(X_0X_3 - X_1X_2) + q(X_2^2 + X_3^2) = 1. \tag{7}$$

This form is positive, being derived from $\sum t_i^2$, and is of determinant $(p^2)^2/p^4 = 1$, hence equivalent to $\sum x_i^2$. Hence (7) has eight solutions $(X_0, \cdots, X_3)$. If t is any of the corresponding values (6), its left-associates exhaust the eight possibilities. Theorem 1 follows.

COROLLARY 1. *Theorem* 1 *holds with m even* [change of notation momentary], *provided v is actually proper and Nv/m is odd.*

For, if $v = ut$ and u is determined up to a right unit factor, then t is determined up to a left unit factor.

COROLLARY 1′. *If z and xz are both proper* (mod m), *and $m \mid Nz$, then z and xz have the same right divisors of norm m.*

3. We now consider the left-multiples of a proper quaternion t.

LEMMA 5. *Let t be proper, $p \nmid t_0^2 + t_\alpha^2$ for some α. No two of the following p^{2k} quaternions are congruent* (mod p^k):

$$(e + fi_\alpha)t, \qquad e, f = 0, 1, \cdots, p^k - 1. \tag{8}$$

From $(r + si_\alpha)t \equiv 0 \pmod{p^k}$ follows $r \equiv s \equiv 0$, for

$$rt_0 - st_\alpha \equiv 0, \qquad rt_\alpha + st_0 \equiv 0, \qquad r(t_0^2 + t_\alpha^2) \equiv 0 \equiv s(t_0^2 + t_\alpha^2).$$

THEOREM 2. *Let t be proper, $p^k \mid Nt$. Then ut represents precisely p^{2k} residues* (mod p^k), *each residue for p^{2k} residues u* (mod p^k).

We cannot have $p \mid t_0^2+t_\alpha^2$ for $\alpha=1, 2$, and 3; for then $p \mid \sum(t_0^2+t_\alpha^2)=2t_0^2+Nt$, $p \mid t_i$ $(i=0, 1, 2, 3)$. Let κ denote the number of solutions x of $xt \equiv 0$, and ρ the number of residues $ut \pmod{p^k}$. The number of solutions x of $xt \equiv wt$, for a given w, is the same as that of $(x-w)t \equiv 0$, hence equals κ; that is, every residue ut is represented for κ residues u, whence $\kappa\rho = p^{4k}$. By Lemma 5, $\rho \geqq p^{2k}$; since $x=(e+fi_\alpha)\bar{t}$ obviously satisfies $xt \equiv 0$, $\kappa \geqq p^{2k}$. Hence $\kappa = \rho = p^{2k}$.

COROLLARY 2. *If $p^k \mid Nt$ the residues* (8) *in Lemma* 5 *represent a complete set of left-multiples of t,* (mod p^k).

COROLLARY 3. *In Theorem* 2 *precisely p^k of the p^{2k} left-multiples of t* (mod p^k) *are pure* (mod p^k).

THEOREM 3. *Let t be proper, $m \mid Nt$. Then all left-multiples ut* (mod m) *which are pure* (mod m) *form an unique set $\mathfrak{M}$; that is, all pure left-multiples of t are proportional* (mod m).

For, by Corollary 3 and the Chinese remainder theorem applied with $m=\prod p^r$, there are precisely m pure left-multiple residues ut (mod m). At least one, say v, is proper, by Lemma 4. Then $0, v, 2v, \cdots, (m-1)v$ exhaust the m possibilities.

THEOREM 4. *To every set $\mathfrak{Q}$ of norm m corresponds one and only one set $\mathfrak{M}$* (mod m), *and conversely, such that $\mathfrak{M}$ contains the pure left-multiples* (mod m) *of the quaternions in $\mathfrak{Q}$, and $\mathfrak{Q}$ contains the right divisors of norm m of the proper elements of $\mathfrak{M}$.*

By taking conjugates we have a similar result for pure right-multiples and left divisors. Since the various left-multiple sets $\mathfrak{M}$ (mod p^r) of x and y are the same and combine, by the C.r.t., into an unique $\mathfrak{M}$ (mod m) we have

COROLLARY 4. *Let x and y be proper* (mod m), $Nx \equiv Ny \equiv 0$. *If x and y have the same right divisors of norm p^r, for every p^r dividing m, then x and y have the same right divisors of norm m.*

COROLLARY 5. *Let v be proper and pure* (mod m), $m \mid Nv$, *and let wv be pure* (mod m). *Then there exists an integer λ such that $wv \equiv \lambda v$.*

4. Conditions for quaternions to have the same divisors will be discussed.

LEMMA 6. *The largest rational integer factor of m dividing z is not changed if z is replaced by uz, or zu, where Nu is prime to m.*

For let $k \mid m$. If $k \mid z$, $k \mid uz$. If $k \mid uz$, $k \mid \bar{u}uz$, $\nmid z$.

LEMMA 7. *If $t', t'', \cdots, t^{(f)}$ are quaternions of odd prime norm p, then $t=t't'' \cdots t^{(f)}$ is proper if and only if*

$$t^{(g)}t^{(g+1)} \text{ is proper for } g = 1, 2, \cdots, f-1. \tag{9}$$

The lemma is trivial if $f=2$. The necessity of (9) is obvious for every f. Assume for a given f that (9) holds and t is proper. Consider $x=tt^{(f+1)}$. If x is not proper, $x=yp$, y integral; and since $p=\bar{t}^{(f+1)}t^{(f+1)}$, $t=y\bar{t}^{(f+1)}$. Hence $t^{(f)}=\theta\bar{t}^{(f+1)}$, since both are right divisors of t of norm p, θ denoting a unit ± 1 or $\pm i_a$. Thus $t^{(f)}t^{(f+1)}=\theta p$.

By a similar argument we obtain

LEMMA 8. *Let $u', u'', \cdots, u^{(h)}$ be of norm p. If any $u^{(\sigma)}u^{(\sigma+1)}$ is improper, it is of the form θp, θ a unit; we can remove the factor p from $u=u'u''\cdots u^{(h)}$, absorb the unit θ into $u^{(\sigma-1)}$ or $u^{(\sigma+2)}$, and proceed with the remaining product of $h-2$ factors. We obtain finally $u=p^r t'\cdots t^{(f)}, t'\cdots t^{(f)}$ proper as in Lemma 7, $f=h-2r$.*

LEMMA 9. *If x and y are proper* (mod p), *and $xy\equiv 0$* (mod p^r), *then $p^r \mid Nx$ and Ny, and x and $\bar{y}$ have the same right divisors of norm p^r.*

For, if $p^r \mid xy$, $p^r \mid \bar{x}xy$ and $xy\bar{y}$, whence $p^r \mid Nx$ and Ny. Hence we can write $x=uu't$, $\bar{y}=\bar{v}\bar{v}'\bar{t}'$, where $Nt=Nt'=p^r$, $Nu'=p^e$, $Nv'=p^f$ $(e,\ f\geqq 0)$, and $p\nmid NuNv$. Then $xy=uu'tt'v'v\equiv 0$ (mod p^r), $u't$ and $t'v'$ being proper. By Lemmas 6 and 8, $tt'\equiv 0$ (mod p^r), that is, t and $\bar{t}'$ are left-associates.

THEOREM 5. *Let x and y be proper* (mod m). *Then x and y have the same right divisors of norm m, if and only if $x\bar{y}\equiv 0$* (mod m).

For if $x=ut$, $y=vt$, and $Nt=m$, then $x\bar{y}=ut\bar{t}\bar{v}=mu\bar{v}\equiv 0$. Conversely, if $x\bar{y}\equiv 0$, we apply Lemma 9 and Corollary 4.

THEOREM 6. *Let x and y be proper* (mod m), *$m \mid Nx$ and Ny. Then there exists a factorization $m=m_1m_2$ in odd positive integers, such that x and y have the same left divisors of norm m_1 and the same right divisors of norm m_2, if and only if*

$$m \mid (x_0y_0 + x_1y_1 + x_2y_2 + x_3y_3). \tag{10}$$

Necessity. If t' and t'' are left and right divisors of x, of norms m_1 and m_2 respectively, then $x=ut''$ where $m_1 \mid Nu$, and the left divisor of norm m_1 of u must be t' (by the uniqueness feature of Theorem 1). Hence we can set $x=t'at''$, $y=t'bt''$, $Nt'=m_1$, $Nt''=m_2$. The expression in (10) is

$$\begin{aligned}\tfrac{1}{2}(x\bar{y}+y\bar{x}) &= \tfrac{1}{2}(t'at''\bar{t}''\bar{b}\bar{t}' + t'bt''\bar{t}''\bar{a}\bar{t}')\\ &= \tfrac{1}{2}Nt''(t'a\bar{b}\bar{t}' + t'b\bar{a}\bar{t}') = \tfrac{1}{2}Nt''[t'(a\bar{b}+b\bar{a})\bar{t}']\\ &= \tfrac{1}{2}Nt''(a\bar{b}+b\bar{a})Nt' = m_1m_2\sum a_ib_i,\end{aligned}$$

Nt'' and $a\bar{b}+b\bar{a}$ being scalars.

Sufficiency. From $x\bar{y}+y\bar{x}\equiv 0$ and $\bar{x}x\equiv 0$ follow

$$x\bar{y}x + y\bar{x}x \equiv 0, \qquad x\bar{y}x \equiv 0 \pmod{m}.$$

Let p^r be any prime-power dividing m. Then $p^r \mid x\bar{y}x$. Let p^s be the highest power of p for which $p^s \mid x\bar{y}$. By Theorem 5, x and y have the same right divisor of norm p^s. If $s \geqq r$ this result suffices as regards p. If $r > s$, we can set $x = ut$, $y = vt$, $Nt = p^s$, whence $u\bar{v}$ is proper and $p^{r-s} \mid Nv$. Then $p^{r-s} \mid u\bar{v}x$. Hence by Theorem 5, $v\bar{u}$ and x have the same left divisors of norm p^{r-s}. These must coincide with the left divisors of v with norm p^{r-s}, and hence with those of y. Corollary 4 now gives us Theorem 6.

The expansion of $x\bar{y} \equiv 0 \pmod{p^r}$ is as follows:

$$x_0y_0 + x_1y_1 + x_2y_2 + x_3y_3 \equiv 0 \pmod{p^r}, \tag{11}$$

$$\begin{cases} (x_0y_1 - x_1y_0) + (x_2y_3 - x_3y_2) \equiv 0 \pmod{p^r}, \\ (x_0y_2 - x_2y_0) + (x_3y_1 - x_1y_3) \equiv 0 \pmod{p^r}, \\ (x_0y_3 - x_3y_0) + (x_1y_2 - x_2y_1) \equiv 0 \pmod{p^r}. \end{cases} \tag{12}$$

THEOREM 5′. *Let $m \mid Ny$, y proper* (mod m). *Then the right divisors of y with norm m are right divisors of x, if and only if, for each prime-power p^r dividing m,* (11) *holds along with* (12_α), *where α ($=1, 2,$ or 3) is such that $p \nmid y_0^2 + y_\alpha^2$. Similarly for left divisors with the $+$'s in* (12) *changed to $-$'s.*

Note that $p \nmid y_0^2 + y_\alpha^2$ for $\alpha = 1$, 2, or 3, since $p \mid Ny$ but $p \nmid y$.

We may assume $\alpha = 1$. Since $p^r \mid Ny$ we readily verify that

$$(-y_0y_2 + y_1y_3)(11) - (y_0y_3 + y_1y_2)(12_1) + (y_0^2 + y_1^2)(12_2) \equiv 0,$$
$$-(y_0y_3 + y_1y_2)(11) + (y_0y_2 - y_1y_3)(12_1) + (y_0^2 + y_1^2)(12_3) \equiv 0.$$

Hence (11)–(12_α) imply (12), and the four congruences (11)–(12) have p^{2r} solutions x (mod p^r), m^2 solutions x (mod m). But by Theorem 2, if $y = vt$, $Nt = m$, then $x = ut$ has m^2 residues (mod m). Every such residue satisfies $x\bar{y} \equiv 0$. Incidentally, this gives an alternative proof of Theorem 5.

COROLLARY 6. *Let x be pure and proper* (mod m), $d^2 + Nx \equiv 0 \equiv e^2 + Nx$. *Then $d + x$ and $e + x$ have the same right divisors of norm m if and only if $d \equiv e$.*

COROLLARY 7. *If x is pure, and v is pure and proper* (mod m), $m \mid Nv$ *and $m \mid \sum x_\alpha v_\alpha$, then there exists an integer x_0 such that $x_0 + x$ and v have the same right divisors of norm m.*

For we have only to choose x_0 to satisfy any one of

$$x_0v_1 + x_2v_3 - x_3v_2 \equiv 0, \quad x_0v_2 + x_3v_1 - x_1v_3 \equiv 0, \quad x_0v_3 + x_1v_2 - x_2v_1 \equiv 0 \ (p^r),$$

in which $p \nmid v_\alpha$, for each prime-power dividing m.

COROLLARY 8. *If y is proper* (mod m), *and $m \mid Ny$, then x has the right divisors of norm m of y, and the left divisors of norm m of y, if and only if $x \equiv ky$* (mod m) *for some integer k.*

For then $x_f y_g - x_g y_f \equiv 0 \pmod{p^r}$, $f, g = 0, 1, 2, 3$.

The value of k may sometimes be obtained from the identity

$$tat = 2(a_0 t_0 - \sum a_\alpha t_\alpha)t - \bar{a}Nt.$$

COROLLARY 9. *If $m \mid Nx$, Ny, and $\sum x_i y_i$, and x and y are proper* (mod m), *then there is a factorization $m = m_1 m_2$ such that*

$$(x_0y_1 - x_1y_0) \equiv \pm (x_2y_3 - x_3y_2), \qquad (x_0y_2 - x_2y_0) \equiv \pm (x_3y_1 - x_1y_3),$$
$$(x_0y_3 - x_3y_0) \equiv \pm (x_1y_2 - x_2y_1),$$

with all the signs $\pm$ taken as $+$ for modulus m_1, and $-$ for modulus m_2.

COROLLARY 10. *If m has no square factor greater than* 1, *x and y are pure and proper, $m \mid Nx$, Ny, and $\sum x_\alpha y_\alpha$, then x and y are proportional* (mod m).

The most interesting special case of Theorem 6 is

THEOREM 7. *Let x and y be proper* (mod p), *p an odd prime dividing Nx and Ny. Then x and y have either the same right divisors or the same left divisors, or both, of norm p, if and only if*

$$x_0y_0 + x_1y_1 + x_2y_2 + x_3y_3 \equiv 0 \pmod{p}. \tag{13}$$

An independent proof involves interesting lemmas:

LEMMA 10. *If $Nt = p$, p cannot divide two of $t_0^2 + t_\alpha^2$ ($\alpha = 1, 2, 3$).*

For if $p \mid t_0^2 + t_\alpha^2 \leqq p$, then $t_0^2 + t_\alpha^2 = p$.

LEMMA 11. *If $Nt = p = Nt'$, then exactly p of the p^2 left-multiples of t,* (mod p), *are also right-multiples of t'.*

For by Lemma 10 we can assume $p \nmid t_0^2 + t_1^2$ and $p \nmid t_0'^2 + t_1'^2$. By Theorem 5′, t' and $(e + fi_1)t$ (cf. Lemma 5) have the same left divisors of norm p if and only if, to modulus p,

$$e(t_0t_0' + t_1t_1' + t_2t_2' + t_3t_3') + f(t_0t_1' - t_1t_0' + t_2t_3' - t_3t_2') \equiv 0,$$
$$e(t_0t_1' - t_1t_0' - t_2t_3' + t_3t_2') + f(-t_0t_0' - t_1t_1' + t_2t_2' + t_3t_3') \equiv 0.$$

The determinant is easily seen to be zero (mod p) while not all the coefficients are zero. Hence there are p solutions e, f.

LEMMA 12. *For a given y such that $p \nmid y$ but $p \mid Ny$, there are precisely $2p^2 - p$ solutions x_0, x_1, x_2, x_3 of*

$$(14) \qquad y_0x_0 + y_1x_1 + y_2x_2 + y_3x_3 \equiv 0, \qquad x_0^2 + x_1^2 + x_2^2 + x_3^2 \equiv 0 \pmod{p}.$$

Although a direct proof is not difficult, we deduce this from a general investigation of such congruences.*

The solutions x of (14) must coincide with the $2p^2-p$ residues having the same right or left divisors of norm p as y, since those residues satisfy (14).

5. Including the original t, there are twenty-four (not necessarily distinct) values $t_0 \pm i_1t_\alpha \pm i_2t_\beta \pm i_3t_\gamma$ in a set $\mathfrak{E}$, obtained by an even number of interchanges and sign-changes of t_1, t_2, t_3. In this way we find a value t in every subset $\mathfrak{Q}$ of $\mathfrak{E}$. We can easily verify

LEMMA 13. *The twenty-four notationally distinct elements in a set $\mathfrak{E}$, with fixed t_0, can be expressed by*

$$\eta t\bar{\eta},$$

$$(15) \qquad \eta = 1,\ i_\alpha,\ (1+i_\alpha)/2^{1/2},\ (1-i_\alpha)/2^{1/2},\ (i_\alpha - i_{\alpha+1})/2^{1/2},\ (i_\alpha + i_{\alpha+1})/2^{1/2},$$

$$\tfrac{1}{2}(1 + i_1 + i_2 + i_3) = \nu,\ i_\alpha\nu,\ \bar{\nu},\ \bar{\nu}i_\alpha,$$

where $\alpha = 1, 2, 3$, and $i_4 = i_1$.

Consider the relation between divisors of elements x and x' in the same set $\mathfrak{E}$. We choose the left-associate of x' having the same real part as x. This left-associate has the same right divisors as x', and is of the form $\eta x\bar{\eta}$, as in (15). From $x = ut$ follows

$$(16) \qquad \eta x\bar{\eta} = \eta u\bar{\eta} \cdot \eta t\bar{\eta}.$$

This is also evident from the fact that an even number of interchanges and sign-changes of i_1, i_2, i_3 produces a simply isomorphic system of quaternions.

The factorization of the quaternions obtained from x by an odd number of permutations and sign-changes of the x_i reduces similarly to that of $\bar{x}$. Trivially, $x = ut$ implies $\bar{x} = \bar{t}\bar{u}$. But the right divisors of x and $\bar{x}$ are not related in any obvious way. Some light on this question is exhibited by

THEOREM 8. *Let x be proper* (mod m), $m \mid Nx$. *Then x and $\bar{x}$ have their right divisors of norm m in the same set $\mathfrak{E}$ if and only if*

$$(17) \qquad m \text{ divides one of } x_f,\ x_f \pm x_g\ (f \neq g),\ x_0 \pm x_1 \pm x_2 \pm x_3.$$

Set $\theta = e\bar{\eta}$, η as in (15), with $e = 1, 2^{1/2}$, or 2 in the respective cases, so that θ is integral, proper, and of norm e^2. Consider

$$(18) \qquad 2s_0 = x\theta + \bar{\theta}\bar{x}, \qquad s_0 \text{ the scalar part of } x\theta.$$

* R. E. O'Connor, *Quadratic and linear congruence*, Bulletin of the American Mathematical Society, vol. 45 (1939), pp. 792–798.

The values of s_0 are precisely the expressions in (17), or their negatives. Let $x=ut$, $Nt=m$. We shall see that

(19) $\bar{x}$ has the right divisor $t'=\eta t\bar{\eta}$, if and only if $m \mid s_0$.

From (18) we obtain

$$e^2\bar{x} = 2\theta s_0 - \theta x\theta = 2\theta s_0 - \theta u\theta\eta t\bar{\eta} = 2\theta s_0 - u't'. \tag{20}$$

If $\bar{x}=vt'$, then $2e^2s_0=yt'$, $m=Nt' \mid 2e^2s_0\bar{t}'$, hence $m \mid s_0$. Conversely, if $m \mid s_0$, $t'=\eta t\bar{\eta}$ is a right divisor of $2\theta s_0$, hence by (20) of $e^2\bar{x}$, hence of $\bar{x}$ (e^2 being prime to m).

THEOREM 9. *Let t be proper, $Nt=m$. Let $\mathfrak{M}_1$ and $\mathfrak{M}_2$ denote, respectively, the set of pure left- and right-multiples of t* (mod m). *Then $\mathfrak{M}_1$ and $\mathfrak{M}_2$ can be obtained from each other by interchanges or sign-changes of i_1, i_2, i_3, if and only if*

$$\text{two of } t_0^2,\ t_1^2,\ t_2^2,\ t_3^2,\ 0 \text{ are equal.} \tag{21}$$

Necessity. Let $v=at$, $v'=tb$, where v and v' are pure (mod m), and v' is obtained from v by permuting or changing signs of v_1, v_2, v_3. Then $\pm v'=\eta v\bar{\eta} = \eta a\bar{\eta}\cdot\eta t\bar{\eta}$, while $-v'=\bar{v}'=\bar{b}\bar{t}$. Since v' is pure, (17) holds trivially with $v=x$, so that $\bar{t}$ and $\eta t\bar{\eta}$ are in the same $\mathfrak{E}$, whence (21).

Sufficiency. If (21) holds, $\bar{t}$ and t are in the same $\mathfrak{E}$. By (16), the left-multiples of t are obtained from those of $\bar{t}$ by permuting and changing signs of i_1, i_2, i_3. The pure left-multiples of $\bar{t}$ are right-multiples of t.

6. Let $h(-n)$ denote the number of classes of primitive, positive, binary quadratic forms $\phi=[k, 2x_0, l]$ of determinant $-n=x_0^2-kl$, and $r'(n)$ the number of proper pure quaternions x of norm n. Gauss showed that, if $n>1$, $n\neq 4f$ or $8f+7$,

$$\begin{aligned} r'(n) &= 12h(-n), \quad \text{if} \quad n \equiv 1 \text{ or } 2 \pmod 4, \\ &= 8h(-n), \quad \text{if} \quad n \equiv 3 \pmod 8. \end{aligned} \tag{22}$$

Assuming that $r'(n)>0$ we demonstrate (22) by means of quaternions.

Let $[x]$ denote the set of four quaternions

$$x, \quad -i_1xi_1 = i_1x_1 - i_2x_2 - i_3x_3, \quad -i_2xi_2, \quad -i_3xi_3, \tag{23}$$

obtained by changing signs of two x_a; four, that is, except for $Nx=1$, when only two of them are distinct. With ϕ we associate the following *process* whereby: (a) *every proper* $[x]$ *of norm n is carried into a certain proper* $[y]$ *of norm n*; (b) *no two distinct proper* $[x]$*'s are carried into the same* $[y]$.

Set $x_0+x=ut$, $Nu=l$, $Nt=k$, which confines t to a set Ω (Corollary 1). Define $y=(tx\bar{t})/k=tu-x_0$. If t is replaced by a left-associate $i_\beta t$, y is replaced

by $-i_\beta y i_\beta$, in $[y]$. If t is changed to $-t$, y is unaltered. If x is replaced by $x'=-i_\alpha x i_\alpha$, $x_0+x'=-i_\alpha u t i_\alpha$, and we again obtain y.

Further, y is proper. For if $p\mid y$, then $p\mid n=Ny$; and $p\mid k$, since $kx=\bar{t}yt$ and x is proper. Since $x_0^2+n=kl$, $p\mid x_0$. Hence $p\mid tu=y+x_0$. Since t is proper, $p\mid tNu$, $p\mid Nu=l$, contradicting the primitivity of $[k, 2x_0, l]$.

Reciprocally, since $\bar{y}=-y$, $-x_0+y=-\bar{u}\bar{t}$, whence $\bar{t}yt=kx$, and the process associated with $[k, -2x_0, l]$ carries $[y]$ back into $[x]$; hence (b).

LEMMA 14. *The process associated with the primitive form $\phi=[k, 2x_0, l]$ is the same as that for the following forms equivalent to ϕ:*

$$[k, 2(x_0+hk), \cdots], \qquad [l, -2x_0, k]. \tag{24}$$

For, if $x_0+x=ut$, then $x_0+hk+x=(u+h\bar{t})t$; and $-x_0+x=-\bar{t}\bar{u}$, $(\bar{u}xu)/l=y=tu-x_0$, since $\bar{y}=-y$.

COROLLARY 11. *Any two equivalent forms ϕ determine the same process.*

LEMMA 15. *Let C, D denote primitive, positive classes of determinant $-n$, CD the product-class under composition. If C carries the proper $[x]$ into $[y]$, and D carries $[y]$ into $[z]$, then CD carries $[x]$ into $[z]$.*

We can choose representative forms in the classes C and D of the types $\phi=[k, 2x_0, hl]$, $\psi=[h, 2x_0, kl]$. By assumption,

$$\begin{aligned} x_0+x&=ut, \quad & Nt&=k, \quad & (tx\bar{t})/k&=y=tu-x_0,\\ x_0+y&=vt', \quad & Nt'&=h, \quad & (t'y\bar{t}')/h&=z=t'v-x_0. \end{aligned}$$

Hence $tu=vt'$; since $Nu=hl$, $u=wt'$,

$$x_0+x=w(t't), \qquad N(t't)=hk, \qquad (t'tx\bar{t}\bar{t}')/(hk)=z.$$

By $\mathfrak{R}=\mathfrak{R}(y)$ we mean the set of pure quaternions obtained from a pure y by permuting and changing signs of y_1, y_2, y_3.

LEMMA 16. *If $Nv=2^r$, then $(\bar{v}yv)/Nv$ is in $\mathfrak{R}(y)$.*

For v is a product of factors $\pm i_\alpha$, $1+i_\beta$. Now $-i_\alpha y i_\alpha$ is in $[y]$; and $(1+i_1)y(1-i_1)=2(i_1y_1-i_2y_3+i_3y_2)$.

LEMMA 17. *If x and y are pure and proper, and $Nx=Ny$, there exists a proper quaternion t of odd norm m such that, for some y' in $\mathfrak{R}(y)$,*

$$txt'=my'. \tag{25}$$

For, $x^2=y^2$, $(x+y)x=y(x+y)$, $(x+y)x(\overline{x+y})=yN(x+y)$. If $y=-x$, the theorem is trivial: $y'=x$, $t=1$. Hence we can suppose $x+y\neq 0$, $x+y=vt$, Nv a power of 2, Nt odd; $tx\bar{t}=m(\bar{v}yv/Nv)$.

If a proper $[x]$ can be carried into $[y]$ by the process associated with a

primitive class C, then $tx\bar{t}=(Nt)y$ must be solvable with t proper and Nt prime to any assignable number. For C contains such forms $[Nt, 2x_0, l]$.

LEMMA 18. *If x is pure, t proper, and $Nt=m$ odd, and if $tx\bar{t}=my$, then we can find an integer x_0 such that $x_0+x=ut$ (whence $y=tu-x_0$).*

For $tx=yt$ has the right and left divisors of norm m of t. By Corollary 8, $tx\equiv x_0 t \pmod m$ for some integer x_0. Hence

$$(x_0+x)\bar{t}=x_0\bar{t}-\bar{x}\bar{t}\equiv x_0\bar{t}-x_0\bar{t}\equiv 0,\quad (x_0+x)\bar{t}=mu,\quad x_0+x=ut.$$

LEMMA 19. *There is at most one primitive class of determinant $-n$ carrying any given proper $[x]$ of norm n into any given proper $[y]$ of norm n.*

For if C and D carry $[x]$ into $[y]$, then both $F=CD^{-1}$ and the principal class E carry $[x]$ into $[x]$. From $tx\bar{t}=mx$, $Nt=m$, follows $tx=xt$, or

$$x_2t_3=x_3t_2,\qquad x_3t_1=x_1t_3,\qquad x_1t_2=x_2t_1.$$

Since x is proper, $t_\alpha=gx_\alpha$ for some integer g, $(\alpha=1, 2, 3)$. The condition $Nt=m$ becomes $t_0^2+ng^2=m$. Thus E represents every integer represented properly by F, $E\sim F$.

LEMMA 20. *Let x' be obtained from x by r sign-changes and s interchanges of x_1, x_2, x_3. Then a proper quaternion t of odd norm such that*

$$tx\bar{t}=(Nt)x' \tag{26}$$

exists if and only if: (α) $x'\equiv x$ (2), (β) *r is even if* $Nx\equiv 3$ (8).

On expanding $tx\bar{t}$ we find $tx\bar{t}\equiv(Nt)x \pmod 2$, whence (α) is necessary. To prove the necessity of (β) we have to prove the impossiblity of $tx\bar{t}=-mx$ with Nt odd, that is, of $tx=-xt$, Nt odd, x_1, x_2, x_3 odd. Expanded, these imply

$$x_1t_1+x_2t_2+x_3t_3=0,\qquad x_1t_0=x_2t_0=x_3t_0=0,$$

whence $t_0=0$, $t_1+t_2+t_3$ is even, Nt is even.

Conversely it suffices to exhibit a solution t of (26) in the cases

$$(\gamma)\ x'=-x,\ Nx\equiv 1 \text{ or } 2 \pmod 4;\quad (\delta)\ x'=i_1x_2+i_2x_1+i_3x_3,\ x_1\equiv x_2\,(2).$$

For (γ), $(x_2i_3-x_3i_2)x(x_3i_2-x_2i_3)=(x_2^2+x_3^2)(-x)$ suffices, since one of $x_2^2+x_3^2$, $x_3^2+x_1^2$ is odd. For (δ), $t=(x_1+x_2)(i_1+i_2)/2+i_3x_3$ is effective, and is of odd norm unless $n\equiv 2 \pmod 4$, in which case $0+x=-(i_1+i_2)t$, where $\bar{u}=i_1+i_2$ (and hence t by Lemma 14) carries x into $i_1x_2+i_2x_1-i_3x_3$, and the sign-change is seen to in (γ). By (α) and (β) we have

COROLLARY 12. *If $Nx\equiv 1$ or 2 (mod 4), x can be transformed into precisely one-third of the vectors in $\mathfrak{R}(x)$. If $Nx\equiv 3 \pmod 8$, x can be transformed into precisely half the vectors in $\mathfrak{R}(x)$.*

Hence the number of primitive classes C is, respectively, one-third and one-half of the number of proper sets $[x]$, and (22) follows.

7. The relations between the representations of n and p^2n as a sum of three squares are easily derived by quaternions.

LEMMA 21. *Every proper, pure y of norm p^2n is of the form $\bar{l}xt$, where $Nt=p$ and $Nx=n$. Here t and x are unique except that they may be replaced by θt and $\theta x\bar{\theta}$, where $\theta=\pm 1$ or $\pm i_\alpha$. If y is changed to $-i_\beta y i_\beta$, we still obtain the same values x. In this way every proper $[y]$ of norm p^2n is derived from an unique proper $[x]$ of norm n.*

For, by Theorem 1, $y=vt$ with $Nt=p$ and t in an unique $\mathfrak{Q}$. Here $p \mid Nv$, whence v has the same left divisors of norm p as $y=-\bar{y}=\bar{l}\bar{v}$. Hence $v=\bar{l}x$, and so on. That x is proper follows from $y=\bar{l}xt$, y proper.

However, a given proper $[x]$ of norm n gives rise to $p-(-n \mid p)$ proper sets $[y]$ of norm p^2n, since there are $p+1$ sets $\mathfrak{Q}$ of norm p and $1+(-n \mid p)$ of them make $\bar{l}xt \equiv 0 \pmod{p}$. The latter values t (by Lemma 18) are the left divisors of x_0+x, where $x_0^2+n \equiv 0 \pmod{p}$. If t and t' are not in the same $\mathfrak{Q}$, $\bar{l}xt$ and $\bar{l}'xt'$ are not in the same $[y]$, by the uniqueness feature of Lemma 21. Hence we have

$$r'(p^2n) = \{p - (-n \mid p)\} r'(n). \tag{27}$$

Using "proper (mod p)" in place of "proper," we obtain

$$r''(p^2n) = \{p - (-n \mid p)\} r''(n),$$

where $r''(n)$ denotes the number of representations of n as a sum of three squares not all divisible by p. Let $r(n)$ denote the total number of representations of n as a sum of three squares. Then

$$\begin{aligned} r(p^2n) &= r(n) + r''(p^2n) = r(n) + \{p - (-n \mid p)\} r''(n) \\ &= r(n) + \{p - (-n \mid p)\} \{r(n) - r(n/p^2)\}, \end{aligned}$$

that is,

$$r(p^2n) = [p + 1 - (-n \mid p)] r(n) - [p - (-n \mid p)] r(n/p^2), \tag{28}$$

for any positive integer n. From this we readily deduce by induction that if h is quadratfrei and $m=\prod p_i^{\alpha_i}$ in powers of distinct primes,

$$\begin{aligned} r(m^2h) &= r(h) \cdot \prod \psi(p_i, \alpha_i; h), \\ \psi(p, \alpha; h) &= \frac{p^{\alpha+1} - 1}{p - 1} - (-h \mid p) \frac{p^\alpha - 1}{p - 1}. \end{aligned} \tag{29}$$

Also, $r'(m^2h) = r'(h) \cdot \prod \phi(p_i, \alpha_i; h)$, where $\phi(p, \alpha; h) = p^\alpha - (-h \mid p) p^{\alpha-1}$.

8. Let h have no square factor greater than 1, m odd and positive. We shall prove that *the general solution of*

$$hm^2 = y_1^2 + y_2^2 + y_3^2 \tag{30}$$

in integers y_α *such that, if* $p \mid y = i_1y_1 + i_2y_2 + i_3y_3$, *then*

$$(-h \mid p) = 1, \quad \text{or} \quad p \mid h \quad \text{and} \quad p^2 \nmid y, \tag{31}$$

is given by $y = txi$, *where*

$$t \text{ is proper and } Nt = m, \ x \text{ is pure and } Nx = h. \tag{32}$$

If $y = txi$, $Ny = Nx(Nt)^2$, whence if (32) holds, y satisfies (30). To see that y also satisfies (31), set $y = p^e z = txi$. Then $p^{2e} \mid Ny = m^2h$, $p^e \mid m = p^e m'$. Hence $txm = p^e zt$, $txm' = zt \equiv \lambda t \pmod m$ for some integer λ by Corollary 8, $m' \mid \lambda = x_0 m'$ since t is proper, $tx \equiv tx_0 \pmod{p^e}$, $p^e \mid t(x_0 - x)(x_0 + x)$, $p^e \mid x_0^2 + h$; $(-h \mid p) = 1$, or $p \mid h$ and $e = 1$.

Conversely we have only to see that every solution y of (30) and (31) is of the required form txi. To do this we write $m = p^r m'$, $r > 0$, $p \nmid m'$, and show that $y = t'zi'$, t' proper of norm p^r, z pure and of norm hm'^2; then no prime p dividing z can satisfy $(-h \mid p) = -1$, or $p \mid h$ and $p^2 \mid z$, since such a prime divides $y = t'zi'$ and contradicts (31). Eliminating in turn each prime-power in m we finally obtain $y = txi$, with $t = t' \cdots t^{(k)}$ proper since the norms of $t', t'', \cdots$ are coprime in pairs. The proof of Lemma 21 extends to

LEMMA 22. *Let* u *be pure,* $p^{2s} \mid Nu$, u *proper* (mod p). *Then* $u = tvi$, t *proper,* $Nt = p^s$, $Nv = (Nu)/p^{2s}$.

Case $p \mid y$, $p \mid h$. By (31), $p^2 \nmid y$. Set $y = pu$, $m = pm_0$, $Nu = km_0^2$. Since u is proper (mod p) and $p^{2r-1} \mid Nu$, $u = tvi$, t proper, $Nt = p^{r-1}$, $Nv = km'^2$. Since $p \mid Nv$, $v = t'w$, $Nt' = p$, and hence $pv = t'v'i'$, where $v' = wt'$. Hence $y = pu = (tt')v'(i'i)$. Here tt' must be proper, since tv (a factor of u) is proper (mod p).

Case $p \mid y$, $(-h \mid p) = 1$. *Set* $y = p^e u$, u proper (mod p), $m = p^e m_0$, $Nu = hm_0^2$. Then $u = tvi$, $Nt = p^{r-e}$, $Nv = hm'^2$. There are two solutions $\pm v_0$ of $v_0^2 \equiv -hm'^2$ (mod p^e). Since $p \nmid v_0$, the left divisors of norm p of $v_0 + v$ and $-v_0 + v$ cannot be the same. Hence we can set $v_0 + v = t'w$, $Nt' = p^e$, t' proper, and choose the sign of v_0 to make tt' proper (Lemma 7). Then $v' = wt' - v_0 = (i'vt')/p^e$, and $p^e v = t'v'i'$, $y = (tt')v'(i'i)$.

Case $p \nmid y$. Then Lemma 22 applies at once.

The restriction that h be quadratfrei can be removed, if the troublesome case where $p^2 \mid h$ and $p^2 \mid y$ is avoided by reducing to h/p^2 and y/p.

9. As an addition to §6 we record the following results which depend only on n and not on the particular x of norm n:

(a) $[1, 0, n]$ carries x into x;

(b) $[2, 2, (n+1)/2]$ carries x into $(x_1, x_3, -x_2)$ if $n \equiv 1 \pmod 4$, $x_2 \equiv x_3 \ (2)$;

(c) $[2, 0, n/2]$ carries x into $(x_1, x_3, -x_2)$ if $n \equiv 2 \pmod 4$, $x_2 \equiv x_3 \ (2)$;

(d) $[4, 2, (n+1)/4]$ carries x into (x_2, x_3, x_1) or (x_3, x_1, x_2) if $n \equiv 3 \ (8)$, depending on the residue of $x_1+x_2+x_3 \pmod 4$, while for the same case,

(e) $[4, -2, (n+1)/4]$ carries x into the other of (x_2, x_3, x_1), (x_3, x_1, x_2).

McGill University,
Montreal, Quebec

A THEOREM ON QUADRATIC FORMS AND ITS APPLICATION IN THE CALCULUS OF VARIATIONS*

BY

M. R. HESTENES AND E. J. McSHANE

1. **Principal result.** Let $P(z), Q_1(z), \cdots, Q_r(z)$ be quadratic forms in the real variables $z_1, \cdots, z_m$ with real coefficients having the following properties: the form $P(z)$ is positive at each point $(z) \neq (0)$ at which the forms $Q_1(z), \cdots, Q_r(z)$ vanish simultaneously; for every set of constants $S_1, \cdots, S_r$, not all zero, the quadratic form $S_iQ_i(z)$† is indefinite; for every linear subspace L on which the quadratic forms $Q_1, \cdots, Q_r$ do not vanish simultaneously, save at $(z) = (0)$, there is a linear combination $S_iQ_i(z)$ of these forms which is positive definite on L. The principal result given in the present note is given in the following:

THEOREM 1. *Under the above hypotheses there exists a set of constants* $S_1, \cdots, S_r$ *such that the quadratic form*

$$P(z, S) = P(z) + S_iQ_i(z) \qquad (i = 1, \cdots, r)$$

is positive definite in the variables $z_1, \cdots, z_m$. *The last two hypotheses may be dropped when* $r = 1$.

This theorem was proposed in a somewhat different form by Bliss in 1937 in a seminar on the calculus of variations. It is useful in sufficiency proofs for multiple integral problems. Proofs of the theorem for the case $r = 1$ were given in the seminar by Albert,‡ Reid,§ McShane and Hestenes, each using a different method. The last two of these proofs have not been published. The proof of Theorem 1 here given is due to McShane and is an immediate extension of the one given by McShane for the case $r = 1$. In the next section Theorem 1 will be applied to the case in which $(z) = (x_1, \cdots, x_n, y_1, \cdots, y_n)$ and the forms $Q_1, \cdots, Q_r$ are the two-rowed minors of the $n \times 2$-dimensional matrix $(x_k\ y_k)$. If $n = 2$ then $r = 1$ and Theorem 2 below is an immediate consequence of Theorem 1. If $n > 2$ the further result described in Theorem 3 below is

* Presented separately to the Society, by McShane on April 8, 1939, and by M. R. Hestenes on December 29, 1939; received by the editors November 29, 1939.

† A repeated index denotes summation.

‡ A. A. Albert, *A quadratic form problem in the calculus of variations*, Bulletin of the American Mathematical Society, vol. 44 (1938), pp. 250–252.

§ W. T. Reid, *A theorem on quadratic forms*, Bulletin of the American Mathematical Society, vol. 44 (1938), pp. 437–440.

needed. This further result for the case $n=3$ has been established by McShane.* The proofs of Theorems 3, 4 and 5 here given are due to Hestenes.†

The proof of Theorem 1 is based on two lemmas, the first of which is the following:

LEMMA A. *Let $m(S)$ be the minimum of the quadratic form $P(z, S)$ on the unit $(m-1)$-sphere $z_kz_k=1$. There exists a set of constants (S_0) which maximizes the function $m(S)$.*

For a set of constants $S_1, \cdots, S_r$ let $M(S)$ be the maximum of the quadratic form $-S_iQ_i$ on the unit sphere $z_kz_k=1$. We have $M(S)>0$ if $(S)\neq(0)$ since the quadratic form $-S_iQ_i$ is indefinite by hypothesis. Moreover $M(hS)=hM(S)$ for every positive number h. Let M_0 be the minimum of $M(S)$ on the set $S_iS_i=1$. Since this minimum is attained, we have $M_0>0$. Finally let N be the maximum of $P(z)$ on the set $z_kz_k=1$. Then for every set of constants (S) there is a point (z) such that $z_kz_k=1$, $-S_iQ_i(z)=M(S)\geqq hM_0$, $S_iS_i=h^2$ and

$$m(S) \leqq P(z, S) \leqq N - M(S) \leqq N - hM_0.$$

It follows that there is a positive constant h_0 such that when $h>h_0$ we have $m(S)<m(0)$ for every set (S) with $S_iS_i=h^2$. There is accordingly a set (S_0) such that the relation $m(S)\leqq m(S_0)$ holds for every set (S), as was to be proved.

LEMMA B. *Let (S_0) be a set of constants which maximizes the function $m(S)$ and set $m_0=m(S_0)$. The set of points (z) satisfying the equation $P(z, S_0)=m_0z_kz_k$ forms a linear space L. There is no quadratic form S_iQ_i which is positive definite on L.*

The set L consists of all points z at which the function $P(z, S_0)-m_0z_kz_k$ attains its minimum value 0. Hence the partial derivatives of that function vanish on L, and therefore the equations

$$P_{z_k}(z, S_0) = 2m_0z_k \tag{1}$$

hold for all z in L. Conversely, if we multiply equations (1) by z_k and sum on k we see that every point z which satisfies (1) lies in L. Thus L is the set of solutions of the linear equations (1), and is therefore linear. Suppose now that there exists a quadratic form $Q=S_iQ_i$ that is positive definite on L. Let K be the unit sphere $z_kz_k=1$ and L_1 the set of points in L on K. Choose $b>0$ such

* E. J. McShane, *The condition of Legendre for double integral problems of the calculus of variations*, Bulletin of the American Mathematical Society, abstract 45-5-209.

† M. R. Hestenes, *A theorem on quadratic forms and its application in the calculus of variations*, Bulletin of the American Mathematical Society, abstract 46-1-83.

that $Q(z)>b$ on L_1, and let N be a neighborhood of L_1 relative to K on which $Q(z)>b$. Let m_1 be the minimum of $P(z, S_0)$ on the closed set $K-N$. Then $m_1>m_0$. It follows that for a sufficiently small positive constant h one will have

$$P(z, S_0 + hS) = P(z, S_0) + hQ > m_0$$

on $K-N$. But $P(z, S_0+hS)>m_0+hb$ on N, and hence $m(S_0+hS)>m(S_0)$, contrary to our choice of the set (S_0). This proves Lemma B.

In order to prove Theorem 1 we note that there is a point $(z)\neq(0)$ on the set L described above at which the forms $Q_1, \cdots, Q_r$ vanish simultaneously. Otherwise by virtue of the last hypothesis made in Theorem 1 there would exist a quadratic form $S_iQ_i(z)$ which is positive definite on L, contrary to Lemma B. At this point (z) we have accordingly $m_0z_kz_k=P(z, S_0)=P(z)>0$ in view of the first hypothesis of the theorem. It follows that $m_0>0$ and hence that $P(z, S_0)$ is positive definite. This proves the first statement in the theorem. The second statement is readily verified. It should be observed that the hypotheses of Theorem 1 imply that there is a point $(z)\neq(0)$ at which the forms $Q_1, \cdots, Q_r$ vanish simultaneously. Otherwise the last hypothesis would imply the existence of a positive definite form S_iQ_i, contrary to the second hypothesis.

The last sentence in Theorem 1 remains to be established. It is easy to see that if $r=1$ the last hypothesis in §1 is automatically satisfied. Suppose then that $r=1$ and that Q is not indefinite. Let K be the sphere $z_kz_k=1$. If Q_1 is positive definite, the sum $P(z, S)$ is positive on K, hence is a positive definite form, provided that S_1 is large. If Q_1 is negative definite, $P(z, S)$ is positive definite provided that $-S_1$ is large. If Q_1 is semi-definite, say positive, let L be the set on which Q_1 vanishes. As in Lemma B, this is linear. On LK the form $P(z)$ is positive; it then remains positive on a neighborhood N of LK relative to K. On $K-N$ the form Q_1 is positive. Choose S_1 large enough so that $P(z, S)$ is positive on $K-N$. On N we have $P(z, S)\geqq P(z)>0$, so that $P(z, S)$ is positive on K, and is therefore a positive definite form.

In the proof of the first part of Theorem 1 made above we have established essentially the following more general result:

COROLLARY. *Suppose the last two of the hypotheses made in Theorem* 1 *hold, and let m be the minimum of the form* $P(z)$ *on the set of points* (z) *satisfying the conditions* $Q_1(z)=\cdots=Q_r(z)=0$, $z_kz_k=1$. *There exists a set of constants* S_i *such that the inequality* $P(z, S)\geqq mz_kz_k$ *holds for all points* (z).

2. **A further result.** Consider now the case in which the space of points (z) described in the last section is of dimension $2n$. For the purposes of this sec-

tion it will be convenient to denote the ith ($i \leqq n$) and the $(n+i)$th coordinates of (z) by x_i and y_i respectively. Thus the points of our space will be denoted by the symbol (x, y). To each point (x, y) there is associated a $n \times 2$-dimensional matrix $(x_i\ y_i)$ whose ith row is composed of the coordinates x_i, y_i of the point (x, y). This matrix will be used below to classify the points (x, y) of our space. By a quadratic form in the variables (x, y) will be meant an expression of the form

$$R_{ik}x_ix_k + S_{ik}x_iy_k + T_{ik}y_iy_k.$$

In particular the expression $S_{ik}x_iy_k$ is a quadratic form in the variables (x, y). Finally by a linear space L of points (x, y) is meant a subspace such that if (x, y) and (x', y') belong to L so also does the point $(ax+bx', ay+by')$, where a and b are arbitrary real constants.

The results described in the last section will be used to prove the following:

THEOREM 2. *Let $P(x, y)$ be a quadratic form in the $2n$ real variables $x_1, \cdots, x_n, y_1, \cdots, y_n$ with real coefficients. Suppose that the inequality $P(x, y) > 0$ holds whenever the $n \times 2$-dimensional matrix $(x_i\ y_i)$ has rank 1. Then there exists an n-rowed skew-symmetric matrix $S = (S_{ik})$ such that the quadratic form $P(x, y) + S_{ik}x_iy_k$ $(i, k = 1, \cdots, n)$ is positive definite.*

To prove this result let Q_{ik} $(i<k)$ be the quadratic form $x_iy_k - x_ky_i$. A linear combination $S_{ik}Q_{ik}$ (summed with $i<k$) is easily seen to be equal to $S_{ik}x_iy_k$ (summed for all i, k) if we set $S_{ii}=0$, $S_{ki}=-S_{ik}$ $(i<k)$. Thus we see that the theorem will be established if we show that the hypotheses of Theorem 1 with Q_i replaced by Q_{ik} are satisfied. The first hypothesis holds since the matrix $(x_i\ y_i) \neq (0\ 0)$ has rank 1 if and only if the forms Q_{ik} vanish simultaneously. Moreover a linear combination $S_{ik}Q_{ik}$ with $S \neq 0$ is indefinite. Finally the last hypothesis of Theorem 1 holds by virtue of the following:

THEOREM 3. *Let L be a linear set of points (x, y) such that the $n \times 2$-dimensional matrix $(x_i\ y_i)$ has rank 2 at each point $(x, y) \neq (0, 0)$ on L. There exists a skew-symmetric matrix $S = (S_{ik})$ such that the quadratic form $S_{ik}x_iy_k$ is positive definite on L.*

Let m be the dimension of L and let $(X_{1\alpha}, \cdots, X_{n\alpha}, Y_{1\alpha}, \cdots, Y_{n\alpha})$ $(\alpha = 1, \cdots, m)$ be a basis for L, that is, a set of m points (X_α, Y_α) in L such that the coordinates of each point (x, y) in L are expressible uniquely in the form

$$x_i = X_{i\alpha}u_\alpha, \qquad y_i = Y_{i\alpha}u_\alpha \qquad (\alpha = 1, \cdots, m). \tag{2}$$

Since the matrix $(x_i\ y_i)$ has rank 2 at each point $(x, y) \neq (0, 0)$ on L, the matrix

$\lambda X+\mu Y$, where $X=(X_{i\alpha})$ and $Y=(Y_{i\alpha})$, has rank m for every pair of real numbers λ, μ, not both zero. Conversely for every pair of matrices X, Y, having $\lambda X+\mu Y$ of rank m when $(\lambda, \mu)\neq(0, 0)$, the corresponding linear space L defined by equations (2) is such that the matrix $(x_i\ y_i)$ has rank 2 at each point $(x, y)\neq(0, 0)$ on L. Moreover by the use of equations (2) it is seen that $S_{ik}x_iy_k=R_{\alpha\beta}u_\alpha u_\beta$, where $R=X'SY$ and X' is the transpose of X. Here and elsewhere it is understood that the symbol for an element of a matrix is obtained by adding a pair of subscripts to the symbol for the matrix. Theorem 3 is accordingly equivalent to the following theorem on matrices:

THEOREM 4. *Let X, Y be two $n\times m$-dimensional matrices such that the matrix $\lambda X+\mu Y$ has rank m for every pair of real numbers λ, μ, not both zero. There exists an n-rowed skew-symmetric matrix S such that the matrix $R=X'SY$ is positive definite, that is, the quadratic form $R_{\alpha\beta}u_\alpha u_\beta$ $(\alpha, \beta=1, \cdots, m)$ is positive definite. Here X' is the transpose of X. The matrix R in general will not be symmetric.*

In order to prove Theorem 4 we first observe that in the proof of the theorem we may replace the matrices X, Y by $X_1=AXB$, $Y_1=AYB$, where A and B are arbitrary nonsingular matrices of dimensions n and m respectively. For, the matrix $\lambda X_1+\mu Y_1$ has rank m when $(\lambda, \mu)\neq(0, 0)$. Moreover, suppose there exists a skew-symmetric matrix S_1 such that the matrix $X_1'S_1Y_1$ is positive definite. Let S be the skew-symmetric matrix $A'S_1A$, and set $R=X'SY$. Then the matrix $B'RB=X_1'S_1Y_1$ is positive definite. It follows readily that R is positive definite and hence that the matrices X, Y can be replaced by X_1, Y_1 respectively. In fact X, Y can be replaced by matrices of the form

$$X=\begin{pmatrix} I \\ O \end{pmatrix}, \qquad Y=\begin{pmatrix} C \\ D \end{pmatrix}, \tag{3}$$

where I is the m-rowed identity matrix, C is an m-rowed matrix and O is the zero matrix. Suppose that X, Y are of this form. Then the condition that the matrix $\lambda X+\mu Y$ be of rank m for all real numbers λ, μ, not both zero, is equivalent to the condition that the matrix

$$\begin{pmatrix} \lambda I-C \\ -D \end{pmatrix} \tag{4}$$

be of rank m for all real values of λ. Moreover the equation $AXB=X$ holds for two nonsingular matrices A and B if and only if A is of the form

$$A=\begin{pmatrix} B^{-1} & B^{-1}E \\ O & F \end{pmatrix}. \tag{5}$$

When A is of this form we have

$$AYB = \begin{pmatrix} B^{-1}(C + ED)B \\ FDB \end{pmatrix}. \tag{6}$$

In the sequel we shall assume that the matrices X, Y, A are of the forms (3) and (5). It will be understood that if $n = m$ the matrices O, D, E, F do not appear in these matrices. Theorem 4 will be established by replacing Y by a matrix AYB having special properties. This will be done with the help of three lemmas, the first of which is the following:

LEMMA 1. *Let $\phi_1(\lambda), \cdots, \phi_k(\lambda)$ be the elementary divisors of the matrix $\lambda I - C$ and let ϵ be an arbitrary constant different from zero. With an elementary divisor of the form $\phi_i(\lambda) = [(\lambda-\alpha_i)^2+\beta_i^2]^{m_i}$ $(\beta_i \neq 0)$ associate the $2m_i$-rowed matrix*

$$M_i = \begin{pmatrix} N_i & O & O & \cdots & O \\ \epsilon I & N_i & O & \cdots & O \\ O & \epsilon I & N_i & \cdots & O \\ \cdot & \cdot & \cdot & \cdots & \cdot \\ O & O & O & \cdots \epsilon I & N_i \end{pmatrix}, \quad N_i = \begin{pmatrix} \alpha_i & -\beta_i \\ \beta_i & \alpha_i \end{pmatrix}, \quad I = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}. \tag{7}$$

With an elementary divisor of the form $\phi_j(\lambda) = (\lambda-\alpha_j)^{m_j}$ associate the matrix

$$M_j = \begin{pmatrix} \alpha_j & 0 & 0 & \cdots & 0 \\ 1 & \alpha_j & 0 & \cdots & 0 \\ 0 & 1 & \alpha_j & \cdots & 0 \\ \cdot & \cdot & \cdot & \cdots & \cdot \\ 0 & 0 & 0 & \cdots 1 & \alpha_j \end{pmatrix}. \tag{8}$$

There exists a nonsingular m-rowed matrix B such that $B^{-1}CB = C_1$, where

$$C_1 = \begin{pmatrix} M_1 & O & O & \cdots O \\ O & M_2 & O & \cdots O \\ O & O & M_3 & \cdots O \\ \cdot & \cdot & \cdot & \cdots \\ O & O & O & \cdots M_k \end{pmatrix}. \tag{9}$$

For it is clear from the special forms (7) and (8) that $\phi_i(\lambda) = |\lambda I - M_i|$. If M_i is of the form (8), the (m_i-1)-rowed minor obtained by deleting the first row and last column has determinant unity. Hence $\phi_i(\lambda)$ is the only non-trivial invariant factor of $\lambda I - M_i$ in this case. Suppose next that M_i is of the

form (7) and let $\phi(\lambda)$ be the determinant of the minor of $\lambda I - M_i$ obtained by deleting the first row and the last column. The values $\phi(\alpha_i+i\beta_i), \phi(\alpha_i-i\beta_i)$ are different from zero since $\epsilon \neq 0$, as one readily verifies. It follows that $\phi(\lambda)$ and $\phi_i(\lambda)$ have no common factors and hence that $\phi_i(\lambda)$ is the only non-trivial invariant factor of $\lambda I - M_i$ when M_i is of the form (7). The matrices $\lambda I - C_1$ and $\lambda I - C$ accordingly have the same elementary divisors and hence the same invariant factors. The matrix C_1 is therefore similar to C, that is, C_1 is of the form $B^{-1}CB$, where B is nonsingular.†

LEMMA 2. *If the equation* $|\lambda I - C| = 0$ *has no real roots, there exists a matrix* C^* *similar to* C *and a skew-symmetric matrix* S *such that the matrix* SC^* *is positive definite. The matrix* SC^* *is not in general symmetric.*

For in this case the diagonal blocks in (9) are of the form (7). Consider C_1 as a function $C_1(\alpha, \beta, \epsilon)$ of the values $\alpha_i, \beta_i, \epsilon$ described in Lemma 1. Let $R(\epsilon) = C_1(0, \beta, \epsilon)$, $S = -R(0)$ and $T = C_1(\alpha, 0, 0)$. Then S and ST are skew-symmetric, S is nonsingular and $C_1 = R(\epsilon) + T$. Since ST is skew-symmetric one has $u_\alpha S_{\alpha\beta} T_{\beta\gamma} u_\gamma = 0$ $(\alpha, \beta, \gamma = 1, \cdots, m)$ and

$$u_\alpha S_{\alpha\beta}(R_{\beta\gamma} + T_{\beta\gamma})u_\gamma = u_\alpha S_{\alpha\beta} R_{\beta\gamma} u_\gamma$$

for arbitrary values of (u). Since $R(0) = -S$ and S is nonsingular, the last quadratic form is positive definite when $\epsilon = 0$ and hence for a value $\epsilon' \neq 0$. The matrices S and $C^* = R(\epsilon') + T$ have the properties described in the lemma.

LEMMA 3. *Let* C, D *be the matrices appearing in the matrix* (4) *and suppose the matrix* (4) *has rank* m *for all real values of* λ. *There exists an* $m \times (n-m)$*-dimensional matrix* E *such that the equation*

$$|\lambda I - C - ED| = 0 \tag{10}$$

has no real roots if the dimension m *of* C *is even and a single real root if* m *is odd.*

Let us begin by disposing of the case $m = n$. If $m = n$, the conclusion of Lemma 3 is that $|\lambda I - C| \neq 0$ for all real λ. This is an immediate consequence of the hypothesis that matrix (4) has rank $m = n$ for all real λ. Incidentally, if $m = n$, both m and n must be even; otherwise the equation $|\lambda I - C| = 0$ would be of odd degree, and would therefore have a real root.

Suppose then that $m < n$. We shall consider first the case in which $m = 2$ and C, D are of the form

$$C = \begin{pmatrix} a & 0 \\ c & b \end{pmatrix}, \qquad D = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \text{ or } D = (d\ e). \tag{11}$$

† A. A. Albert, *Modern Higher Algebra*, University of Chicago Press, 1937, pp. 84–85.

If D is the two-rowed identity matrix, let

$$G = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}, \qquad E = G - C.$$

Then the equation (10) takes the form $\lambda^2+1=0$ and has no real roots. Suppose next that $D=(d\ e)$. We may suppose that $b=0$ since this can be brought about by replacing λ by $\lambda+b$. Since the matrix (4) has rank 2 for $\lambda=0$ and $\lambda=a$, we have $e\neq 0$ and $ad+ec\neq 0$. Choose numbers α, β such that

$$\alpha ce - \beta ae = -1, \qquad \alpha d + \beta e + a = 0.$$

Then equation (10) with $E'=(\alpha\ \beta)$ as the transpose of E reduces to $\lambda^2+1=0$. The lemma is accordingly true for the matrices (11).

To prove the lemma as stated, let h be the number of real roots of the equation $|\lambda I-C|=0$, each root counted a number of times equal to its multiplicity. We may suppose that $h\geqq 2$. It is sufficient to show that the matrix E can be chosen so that the equation (10) has exactly $h-2$ real roots. In view of equation (6) it is sufficient to prove this result when C, D are replaced respectively by matrices of the form $C_1=B^{-1}CB$, $D_1=F DB$, where B and F are nonsingular. By virtue of Lemma 1 we may select B so that C_1 is given by (9). In fact B may be chosen so that the matrices (8) corresponding to real roots of $|\lambda I-C|=0$ have higher subscripts than the matrices (7) corresponding to roots that are not real. It follows that, after a suitable choice of the matrix F in D_1, one has

$$C_1 = \begin{pmatrix} M & O \\ N & C_2 \end{pmatrix}, \qquad D_1 = \begin{pmatrix} U & O \\ V & D_2 \end{pmatrix},$$

where the matrices C_2, D_2 are the matrices C, D in (11). The matrix (4) with $C=C_2$, $D=D_2$ has rank 2, since otherwise the corresponding matrix with $C=C_1$, $D=D_1$ could not have rank m. It follows that there exists a matrix E_2 such that the equation (10) with $C=C_2$, $D=D_2$, $E=E_2$ has no real roots. Choosing

$$E = \begin{pmatrix} O & O \\ O & E_2 \end{pmatrix},$$

the equation (10) with $C=C_1$, $D=D_1$ reduces to the product

$$|\lambda I_{m-2} - M| \cdot |\lambda I_2 - C_2 - E_2 D_2| = 0,$$

where the subscript on I denotes its dimension. This equation has $h-2$ real roots and the proof of Lemma 3 is complete.

We are now in position to prove Theorem 4. As was seen above we can assume that the matrices X, Y are of the form (3). Since we can replace Y

by the matrix (6) we can suppose, by virtue of Lemma 3, that the equation $|\lambda I - C| = 0$ has no real roots if the dimension m of C is even and one real root if m is odd. Consider first the case in which m is even. Then by Lemma 2 there is a nonsingular matrix B and a skew-symmetric matrix S such that the matrix SC^* with $C^* = B^{-1}CB$ is positive definite. Replace Y by AYB, where A is defined by equation (5) with $E = 0$ and $F = I$. Then C is replaced by C^* so that the matrix SC is now positive definite. Let S_1 be the n-rowed skew-symmetric matrix

$$\begin{pmatrix} S & O \\ O & O \end{pmatrix}.$$

The product $X'S_1Y$ is then equal to SC and is accordingly positive definite. This proves Theorem 4 for the case in which m is even.

The case in which m is odd will be reduced to the case in which m is even. As was seen above we can assume that the equation $|\beta I - C| = 0$ has only one real root $\lambda = a$. Choose nonsingular matrices B, F such that the matrices $C_1 = B'CB$, $D_1 = FDB$ are of the form

$$C_1 = \begin{pmatrix} M & O \\ O & a \end{pmatrix}, \qquad D_1 = \begin{pmatrix} U & b \\ V & O \end{pmatrix} \qquad (b \neq 0).$$

This choice is possible by virtue of Lemma 1 with $M_k = (a)$ and the fact that the matrix (4) has rank m when $\lambda = a$. We can suppose that $C = C_1$, $D = D_1$, since this result can be brought about by replacing Y by AYB, where A is given by (5) with $E = O$. Clearly $m < n$ since D has at least one row. Let X_1, Y_1 be the matrices

$$X_1 = \begin{pmatrix} I & O & O \\ O & 1 & 0 \\ O & 0 & 1 \\ O & O & O \end{pmatrix}, \qquad Y_1 = \begin{pmatrix} M & O & O \\ O & a & -b \\ U & b & a \\ V & O & O \end{pmatrix}$$

obtained by adding a suitable column to each of the matrices X and Y. Since the equations $|\lambda I - M| = 0$, $(\lambda - a)^2 + b^2 = 0$ have no real roots, the matrix $\lambda X_1 - Y_1$ corresponding to (4) has rank $m+1$ for all real values of λ. As was seen in the last paragraph, there exists, since $m+1$ is even, a skew-symmetric matrix S such that the matrix $X_1'SY_1$ is positive definite. The matrix $X'SY$, being a principal minor of $X_1'SY_1$, is also positive definite. This completes the proof of Theorem 4 and hence also of Theorem 3.

Incidentally, by an argument similar to that just made, it can be shown that if L is a maximal linear space having the properties described in Theorem 3 then its dimension is n if n is even and $n-1$ if n is odd.

3. **Application to the calculus of variations.** Consider the problem of minimizing the double integral

$$I = \iint_A f(x, y, z_1, \cdots, z_n, p_1, \cdots, p_n, q_1, \cdots, q_n)dxdy,$$

where $p_i = \partial z_i/\partial x$, $q_i = \partial z_i/\partial y$, in a class of subspaces

$$z_i(x, y) \qquad ((x, y) \text{ on } A + B;\ i = 1, \cdots, n) \tag{12}$$

having a common boundary, where B is the boundary of A. The integrand $f(x, y, z, p, q)$ is assumed to be continuous and to have continuous first and second partial derivatives on a region $\mathcal{R}$ of points (x, y, z, p, q). The subspaces (12) are assumed to be continuous, to have continuous second partial derivatives and to have their elements (x, y, z, p, q) in $\mathcal{R}$. The boundary B of A is supposed to be a simply closed continuous curve having a piecewise continuously turning tangent. Weaker differentiability assumptions could be made.

LEMMA 4. *Let $S_{ik}(x, y) = -S_{ki}(x, y)$ $(i, k = 1, \cdots, n)$ be arbitrary continuous functions having continuous first and second derivatives in a neighborhood of the set $A + B$. Then the integral*

$$\begin{aligned} J &= \iint_A g(x, y, z, p, q)dxdy \\ &\equiv \iint_A (1/2)\left\{p_i \frac{\partial}{\partial y}(s_{ik}z_k) - q_i \frac{\partial}{\partial x}(s_{ik}z_k)\right\}dxdy \end{aligned} \tag{13}$$

has the same value for all subspaces (12) *here considered.*

For by virtue of Green's theorem the value of $2J$ is given by the formula

$$2J = \iint_A \left\{\frac{\partial}{\partial y}(S_{ik}p_iz_k) - \frac{\partial}{\partial x}(S_{ik}q_iz_k)\right\}dxdy = \int_B z_kS_{ki}dz_i,$$

and hence is completely determined by the common boundary of the subspaces (12).

Let E be a minimizing subspace. Then E must satisfy the *condition of Legendre*,* that is, at each element (x, y, z, p, q) on E the inequality

$$P(\xi, \eta) = f_{p_ip_k}\xi_i\xi_k + 2f_{p_iq_k}\xi_i\eta_k + f_{q_iq_k}\eta_i\eta_k \geqq 0 \tag{14}$$

must hold for every set (ξ, η) whose $n \times 2$-dimensional matrix has rank 1. If

* Graves, *The Weierstrass condition for multiple integral variation problems*, Duke Mathematical Journal, vol. 5 (1939), pp. 656–660. Graves proves only the Weierstrass condition. The condition of Legendre is a well known consequence of that of Weierstrass.

the condition (14) with the equality excluded holds on E for every set (ξ, η) whose matrix $(\xi_i \eta_i)$ has rank 1, then E will be said to satisfy the *strengthened condition of Legendre.* It should be emphasized that in this strengthened condition of Legendre we require that the inequality $P(\xi, \eta) > 0$ hold only for sets (ξ, η) whose $n \times 2$-dimensional matrix $(\xi_i \eta_i)$ has rank 1 and not for all sets $(\xi, \eta) \neq (0, 0)$. We have the following:

THEOREM 5. *Let E be a subspace* (12) *satisfying the strengthened condition of Legendre. There exists an invariant integral of the form* (13) *such that for the function $F = f + g$ the inequality*

$$F_{p_i p_k} \xi_i \xi_k + 2F_{p_i q_k} \xi_i \eta_k + F_{q_i q_k} \eta_i \eta_k > 0 \tag{15}$$

holds at each element (x, y, z, p, q) on E for every set $(\xi, \eta) \neq (0, 0)$.

For by Theorem 2 there exists for each point on E a skew-symmetric matrix (S_{ik}) such that the quadratic form $P(\xi, \eta) + S_{ik}\xi_i\eta_k$ is positive definite at this point and hence in a neighborhood of this point. It follows readily that the set $A + B$ can be covered by a finite number of circles $C_1, \cdots, C_m$ having a common radius r such that to the circle C_α' of radius $2r$ concentric with the circle C_α there is associated a skew-symmetric matrix (S_{ik}^α) of constant elements such that the form $P_\alpha(\xi, \eta) = P(\xi, \eta) + S_{ik}^\alpha \xi_i \eta_k$ is positive definite at each point on E whose projection in the xy-plane is in C_α'. We propose now to construct a set of functions $S_{ik}(x, y) = -S_{ki}(x, y)$ having continuous second derivatives on a neighborhood of $A + B$ and such that the form $P(\xi, \eta) + S_{ik}(x, y)\xi_i\eta_k$ is positive definite on E. To this end let $h(t)$ be a function having continuous second derivatives and such that $h(t) = 0$ when $t \leqq r$, $h(t) = 1$ when $t \geqq 2r$, and $0 < h(t) < 1$ when $r < t < 2r$. Let (x_α, y_α) be the center of the circle C_α, and let $d_\alpha(x, y)$ be the distance between the points (x, y) and (x_α, y_α). Set $h_\alpha(x, y) = h[d_\alpha(x, y)]$. It should be observed that the product $h_1 h_2 \cdots h_m$ is identically zero on the set $A + B$. This follows because $h_\alpha \equiv 0$ on C_α and the circles $C_1, \cdots, C_m$ cover $A + B$. Denote by $S_{ik}(x, y)$ the function

$$(1 - h_1)S_{ik}^1 + h_1(1 - h_2)S_{ik}^2 + \cdots + h_1 h_2 \cdots h_{m-1}(1 - h_m)S_{ik}^m,$$

where S_{ik}^α $(\alpha = 1, \cdots, m)$ are the constants described above. The quadratic form $P(\xi, \eta) + S_{ik}(x, y)\xi_i\eta_k$ determined by these functions is easily seen to be identical with the sum

$$(1 - h_1)P_1(\xi, \eta) + h_1(1 - h_2)P_2(\xi, \eta) + \cdots + h_1 h_2 \cdots h_{m-1}(1 - h_m)P_m(\xi, \eta),$$

where $P_\alpha(\xi, \eta)$ is the quadratic form associated with the circle C'. Each term in this sum is positive or zero, since for each integer α the product

$(1-h_\alpha)P_\alpha(\xi, \eta)$ is positive definite at points on E corresponding to points in C_2' and identically zero elsewhere. Moreover the terms do not vanish simultaneously when $(\xi, \eta) \neq (0, 0)$. The quadratic form $P(\xi, \eta)+S_{ik}\xi_i\eta_k$ is accordingly positive definite on E.

We now use these functions S_{ik} to define the function g of equation (13). From the definition of g we at once obtain

$$g(x, y, z, p, q) = (1/2)[(\partial S_{ik}/\partial y)p_iz_k + (\partial S_{ki}/\partial x)q_iz_k] + S_{ik}p_iq_k.$$

This shows that the Legendre quadratic form (15) for g is $S_{ik}\xi_i\eta_k$. Hence if $F=f+g$ the Legendre form (15) for F is exactly the positive definite form $P(\xi, \eta)+S_{ik}\xi_i\eta_k$. This establishes Theorem 5.

THE UNIVERSITY OF CHICAGO,
CHICAGO, ILL.,
THE UNIVERSITY OF VIRGINIA,
CHARLOTTESVILLE, VA.